An Investigation of Ensemble Learning in Dust Storm Prediction Using Machine Learning Techniques

Tariq Saad Almurayziq

A thesis submitted in partial fulfilment of the
requirements of the University of Brighton
for the Degree of Doctor of Philosophy

October 2018

Abstract

The main attributes that are known to contribute to dust storms are wind speed, temperature, air pressure, humidity and the type of surface. Weather experts expend considerable effort to achieve a high degree of forecast accuracy when detecting this weather phenomenon and modern technology means that it is now possible to detect dust storms using satellite images. However, modern applications have yet to be applied using historical data for dust events prediction. In another word, most of current forecasting models are build based on satellite monitor, and in proposed approach it argues that the new techniques such as AI could asset in weather forecasting. The archive of historical dust storm events is a valuable database and it is quite possible that applying one or a combination of artificial intelligence (AI) techniques will help to devise a reliable way of predicting future dust events based on these old cases of dust events. This study examines the process of predicting and identifying dust storms by focusing on a Bayesian network (BN) with a case-based reasoning (CBR) approach and rule-based system (RBS).

The aim of this thesis is to examine trends in CBR by exploring some of the challenges and perceived benefits that accrue from the use of CBR in predicting and identifying dust storms. In addition, this study seeks to determine the applicability of CBR and its appropriateness for predicting dust storms. Indeed, some of the previous studies and related findings will form the basis on which the study's evaluation of CBR's validity and applicability will be determined.

The current study's findings, it is arguing that the CBR with other AI techniques could help to predict future dust storm events. The results reveal that there is similarity between dust cases, and there are many attributes of dust storm play more important roles than others in dust prediction such as wind speed important than air temperature, which could provide opportunities for us to search for the optimal weight of these attributes.

The BN-CBR combination illustrates the degree of accuracy in terms of predicting forthcoming dust storms relative to the pure CBR. BN-CBR is able to forecast future dust storms by making comparisons with similar events by reclassify the old dust events using BN and predict coming new dust storm using Nearest neighbour (NN), the ideal value as demonstrated 3NN.

A real example has been used to test the RBS and the results have been positive, satisfying the BN-CBR prediction in the short term. In addition, suitable actions have been delivered that could usefully serve targeted sectors and inform the wider community about

this weather phenomenon. Although this is still in the early stages of being developed, the initial results are encouraging. This suggests that the selected approach could prove useful tool to predict future dust events, by using combinations of AI techniques.

Table of Contents

List of Publications

1. **Towards successful prediction of dust storms using case-based reasoning and artificial neural network, UKCBR 2015**, Tariq Al Murayziq, Kapetanakis, Stylianos and Miltos Petridis, Towards successful prediction of dust storms using case-based reasoning and artificial neural network In: 20th UK Workshop on Case-Based Reasoning, Cambridge, UK, Dec 2015.

2. **Towards successful dust storm prediction using Bayesian network and case-based reasoning, UKCBR** 2016, Tariq Al Murayziq, Kapetanakis, Stylianos and Miltos Petridis, Towards successful prediction of Bayesian network and case-based reasoning In: 21th UK Workshop on Case-Based Reasoning, Cambridge, UK, Dec 2016.

3. **Identifying and predicting the dust events by using case based reasoning (CBR), UKCBR 2017**, Tariq Al Murayziq, Kapetanakis, Stylianos and Miltos Petridis, Identifying and predicting the dust events by using case based reasoning (CBR) In: 22th UK Workshop on Case-Based Reasoning, Cambridge, UK, Dec 2017.

4. **Predicting dust storms using hybrid intelligence system, AI 2017**, Tariq Al Murayziq, Kapetanakis, Stylianos and Miltos Petridis, Predicting Dust Storms Using Hybrid Intelligence System. In Cambridge, UK, Dec 2017.

5. **Intelligent signal processing for dust storm prediction using ensemble case-based reasoning, International Conference on Tools with Artificial Intelligence (ICTAI 2017)**, Tariq Al Murayziq, Kapetanakis, Stylianos and Miltos Petridis, Intelligent Signal processing for Dust Storm prediction using ensemble Case-based Reasoning. In: Boston, US, Nov 2017.

List of Tables

List of Figures

Acknowledgments

This study is devoted to my father Said Ashiq Al Murayziq, my mother Fatimah Mohammed Al Murab and my wife for their understanding and support. I would really like to express my heartfelt thanks to my supervisor Dr Stelios for all of his encouragement, advice and support during the work on this thesis. Also, I would like to thank Dr Miltos for his advice.

Thanks to Dr Abdallah Almisind, the weather expert in the Middle East for the time that he spent discussing all matters regarding weather phenomena in the region.

Finally, I would thank the University of Brighton for the great opportunity to study and research in this area and meet these experts.

Abbreviations and Acronyms

AI	Artificial intelligence
ANN	Artificial neural network
AVHRR	Advanced very high-resolution radiometer
BT	Brightness temperature
BTD	Brightness temperature differences
BN	Bayes network
CV	Cross validation
CBR	Case-based reasoning
RBS	Rule based system
R2	Coefficient of determination
RMSE	Root-mean-square error
TAN	Tree augmented naïve Bayes
TIR	Thermal infrared
IR	Image recognition
ICTAI	International conference on tools with artificial intelligence
KNN	K nearest neighbours
LOOCV	Leave-one-out cross-validation
MB	Markov blanket
MSE	Mean squared error
MAE	Mean absolute error
NWP	Numerical weather prediction
UNEP	United Nations Environment Programme

UKCBR	United Kingdom conference in case based reasoning
WMO	World Meteorological Organisation
WHO	World Health Organisation
UNCCD	United Nations Convention to Combat Desertification

Declaration

I certify that this work has not been accepted in substance for any degree and is not concurrently being submitted for any degree other than that of Doctor of Philosophy (Ph.D.) being studied at the University of Brighton. I also declare that this work is the result of my own investigations except where otherwise identified by references and that I have not plagiarised the work of others.

____

Tariq Almurayziq

Supervisors

____

Dr. Stylianos Kapetanakis

____

Professor Miltos Petridis

Dedicated: *to my family for their incredible encouragement in life.*

Chapter One

Introduction

As the severity and number of dust storms grow, weather experts and researchers evolve in their efforts to understand dust event procedures and produce tools to predict future storms with high accuracy. The behaviour of dust storms varies based on five attributes: temperature, pressure, wind speed, humidity and surface. Dust storms may change a person's lifestyle because of the difficulties of doing outdoor activities in dusty weather. However, the ability to predict dust storm using the history of previous storms has not been mastered yet. This study examines the process of predicting and identifying dust storms by focusing on a Bayesian network (BN) with case-based reasoning (CBR) approach and rule-based system techniques (RBS).

Case-based Reasoning (CBR) is an Artificial Intelligence technique that works in a similar way, by retrieving and reusing knowledge and adapting it to the problem in hand. It is therefore worth investigating its application within the concept of weather forecasting specifically dust storm condition (Kolodner, 1993). In addition, the proposed methodology explains the aim behind combining Bayesian network (BN) with (CBR), come up with reasonable prediction and enhance the prediction accuracy of CBR. Bayesian Networks (BNs) are a member of probabilistic graphical models for modeling uncertainty, and it is a powerful tool for subjective logic explanations (Ben-Gal et al., 2007).

The BN-CBR approach outperformed alternative AI techniques such as decision trees (DT) and artificial neural network (ANN). As such, BN-CBR are better than these alternatives to forecast future dust storms. Given that the dust cases archive is not well classified, the BN will reclassify the whole dust dataset to place each dust event in a proper category based on probability. In this study, three BN structure models have been taken into consideration in order to achieve more realistic forecasts. CBR forecasts future dust storms based on reflections of previous dust storms that fit with the nature of the current study.

A RBS prototype developed to retrieves the BN-CBR prediction and executes the appropriate rules to generate actions that could reduce the adverse impacts of dust events. At this stage weather experts are involved to evaluate and validate dust cases and actions to ensure that each alarm generated from this intelligence combination is accurate and fits with the level of risk. This chapter describes the reasons for choosing CBR, the motivation, aims and objectives of the current study, and the formulated research questions. Subsequent chapters set out precisely how this combination delivers more accurate forecasts of dust storms than the available alternatives.

Defining and Understanding of Dust Storms

A dust storm is a meteorological phenomenon that commonly occurs in desert regions where strong winds blow away the loose sand and dirt from dry surfaces. In this case, the particles are transported through suspension and saltation, which are the processes by which the surface is moved from one place to another. Various researchers and academic scholars have recently claimed that poor management and prevention measures to address the problem of dry lands across the world are the major causes of these occurrences. For instance, negligence of a fallow system can increase the frequency and size of a dust storm, thereby changing both local and global climate, which in turn affects the economic world. When predicting a dust storm, it is appropriate to assess its causes as a way of determining best-forecast methodologies. Among the stated causes of dust storms are thunderstorm outflows, strong pressure gradients that increase wind velocity, wind, drought, poor farming, and grazing (Gillette, 1980; see also Pye, 2015).

Dust storms have been associated with the spread of diseases across the world because virus spores on the ground are blown up into the atmosphere and they then act as acid rain or smog (Vidal, 2009; Aood & Jadd, 2016). Dust can also cause impairment to electronics and human vision. It can also affect the transportation and settlement centres of people living in the affected environments. As such, dust storm prediction helps affected groups and regions to prepare for its occurrence through preventative measures such as taking cover, vacating the streets to avoid accidents, sealing windows and doors, and providing security to outdoor property such as equipment and vehicles. It is also helpful because farmers can have an early harvest, store any farm equipment and secure their livestock.

Rationale

This study is significant because it seeks to examine the trends, obstacles and perceived solutions to the challenge of using CBR to predict dust storms based on previous storms. Some of the government sectors, firms and personnel might benefit from this study because of the data it provides regarding current trends and new tactics that can be used to predict future dust storms. Moreover, other individuals with an interest in meteorology may benefit from the mechanistic analysis that prevents the procedures of predicting future dust events based on just the historical dust cases and without satellite images. What follows below is a brief description of the government sectors and individuals that might benefit from this study:

Health sector: Dust storms could well cause serious health problem such as asthma, sore eyes, runny noses and the spread of viruses (Onishi et al., 2015). The problem is that with current forecasting models, health centres cannot be notified in time for them to be prepared for an influx of patients if a dust storm occurs close to the city.

Education sector: Studies have shown that a number of schools have been closed in Saudi Arabia during the middle of the day due the dust storms (Moradian et al., 2015). Late notification results in sending students home in the middle of the day with associated risks to their health.

Transportation sector: Low vision caused by dust events could lead to unexpected traffic congestion and accidents. In the US, large numbers of accidents have been reported after dust storm events (Crooks et al., 2016).

***Weather experts & farmers*:** Archiving and monitoring dust storm cases is beneficial to weather experts to understand this phenomenon and clarify the reasons. In this study, AI combination techniques will make it possible to archive and report dust cases. In addition, farmers could also benefit and preserve their harvest. Maleki et al. (2017) report that Iran has recorded low agricultural production as a result of dust storms.

Research Questions

Can we effectively predict dust storms using a hybrid intelligence system by combining a Bayesian network with case-based reasoning (CBR) techniques?

CBR systems provide an effective solution to various problems as they occur in real-time (Kapetanakis & Petridis, 2014). CBR techniques have been applied in a wide spectrum of settings and environments to help people make appropriate decisions. Arguably the main advantage of CBR is that it replicates the way in which human experts go about solving problems with reference to historical events that were similar to the situation currently being faced. This is a logical approach to solving problems. Alternative methods such as rule based system (RBS) and artificial neural network (ANN) are associated with a number of known disadvantages and these issues will be discussed in detail in the following chapters.

The empirical literature suggests that the available data along with probabilistic theory provide a suitable theoretical basis for managing uncertainty. This is a problem that often arises with probability weather forecasts. However, it is also apparent that the statistical methods relied upon in the past have been computationally complicated. A more practicable approach would be to apply BNs for statistical solutions when conducting probability weather forecasts (de Kock et al., 2008). Therefore, the current research considers the feasibility of deploying BN-CBR combinations to help forecast future dust events in a timely manner.

Can we identify the key attributes of dust storms using the CBR technique?

It is the main features of dust storms that play a central role in the current study. In order to arrive at accurate results, it is necessary for the features to be clearly specified. There are several ways in which this can be achieved. The way in which the main dust characteristics are established in the current study is by applying the CBR technique. These results are then supported by the algorithm ranker and expert rankings.

Can case-based reasoning effectively assist in generating applied rules and safety actions for expected dust storms?

If a system is to be relied upon to assist with the decisions taken by humans, it is necessary for it to justify why it has made the recommendations that it has. Such explanations help to develop confidence in the predictions that are formed so that appropriate actions can be taken. Importantly, the output generated by CBR systems is transparent and accompanying explanations are provided (Kapetanakis et al., 2010). These benefits are incorporated into the RBS system which sets rules and suggests actions to help mitigate the harmful effects of dust events.

Aims and Objective

The aim of this thesis is to examine trends in case-based reasoning by exploring some of the challenges and perceived benefits that accrue from the use of CBR in predicting dust storms based on historical storms. In addition, this study seeks to determine the applicability of CBR and its appropriateness for predicting future dust events.

Motivation

Dust storms afflict numerous areas of the world, causing health problems, property damage and inconvenience. Saudi Arabia is especially prone to the adverse effects of dust storms and extreme weather conditions. Saudi Arabia is therefore selected as a testbed in this study to confirm the effectiveness of the proposed techniques, so that populations can take evasive action to protect their health and property. Especially with the cities that located near or surrounded by desert.

Thesis Structure

Chapter 2 provides information regarding dust storm events and their impact on both human life and the land. Also, this chapter refers to various studies that have applied CBR for problem-solving and explains how BN is used to solve uncertainty in CBR and enhance the accuracy of predictions. The review also establishes the degree of success when using CBR in the selected studies. Furthermore, the literature review focuses on the challenges of predicting dust storms. Of particular concern is to explore the trends, challenges and benefits that arise from applying CBR in the predictions on which this study focuses. Also, CBR enables the examination of mechanisms that could be applied to predict future dust storms. This will not only serve to minimise the consequences of dust events but also contribute to the general AI knowledge base and prevent adverse effects arising from future weather conditions based on historical experience.

Chapter 3 provides an extensive explanation of how CBR is applied for forecasting dust storms. It is also worth noting that this chapter describes how CBR combines with other approaches such as Bayesian network(s) and the Rule Based System (RBS). And the different between BN node structure, in terms of how the BN structure effect in building the BN model. The CBR approach has been adopted and precedes the data examination, evaluation and integration process. The methodology culminates with analysis and discussion of the results from a specific case and concludes with recommendations regarding the future of using CBR in predicting the weather.

Chapter 4 presents the experiments and findings in terms of using CBR to forecast. It is also shown how BN could be integrated with CBR to increase the accuracy of predictions compared with the alternative AI techniques such as ANN. Also presents the results that have been found throughout this investigation. In addition, this chapter provides details of the outcome of applying RBS in real dust examples.

Chapter 5 discusses how each technique is valid and fits to the whole approach in terms of how complete and appropriate all of the available data and methods are to realise the best possible results. It explains how the outcomes are evaluated in terms of how accurately the results compare to real time cases and how safety actions are simulated in reality.

Chapter 6 concludes the thesis. The contributions made by the study are presented and evaluated. Advice is offered for how the study could be extended by future researchers, including areas of interest that could be explored.

Chapter Two

Literature Review

Definition and Reasons for Dust Storms

Dust storms are defined as the result of surface winds raising large quantities of dust into the air and decreasing visibility at eye level to less than 1000m. Dust storms are events in the atmosphere which result from the corrosion and transportation of mineral residues from the surface (UNEP, WMO & UNCCD, 2016). Dust storms are typically related to arid and semi-arid regions but happen anywhere where there are dry unprotected residues. The dust storm is a severe condition of weather that is characterised by strong winds and also dust-filled air over a widespread area. A dust storm is differentiated from a sand storm on the basis of particle size. Studies have indicated that dust storms are comprised of a multitude of very fine particles, whereas sand storms have bigger particles that vary from .08mm to 1mm (UNEP, WMO & UNCCD, 2016).

Dust storms are caused by strong winds over loose sand and they pick up a lot of that material which means that visibility is greatly reduced. The extensive abundance of loose sand in deserts makes them the most common sites for dust storms to form. Studies have indicated that in desert areas at specific times of the year, dust storms become more common due to strong air heating over the desert, causing the instability of the lower atmosphere. The lower atmosphere's instability combines strong winds in the internal troposphere down to the surface, generating winds that are stronger at the surface (Sissakian et al., 2013). Studies have stated that dust storms are the result of many inter-dependent aspects that span across many regional countries. Many years of unsuitable practices of farming, water resource mismanagement as well as change in climate continues to contribute to the decreased coverage of vegetation, droughts and desertification that directly contribute to the increasing problem of regional dust storms. It is noted that arid regions and drought conditions favour soil particle dissolution and wind contributes to the occurrence of dust storms (Sissakian et al., 2013)

Numerous investigations have showed that dust storms happen as a result of a series of interlinked direct as well as indirect drivers working at diverse scales. These drivers include the following:

Direct Drivers in Natural Ecosystems

Studies have explained that wind is the main driver of dust storms in all systems. Particular synoptic meteorological conditions that generate winds differ in diverse regions. A worldwide review of tendencies in terrestrial near-surface wind speeds shows that reductions in wind speed are widespread geographically in recent decades, with drops being indicated in the mid-latitudes and topics of both hemispheres but with increases indicated at high-latitudes. It is vital to note that lower wind speeds at low latitudes are expected to reduce the risk of dust storms. The action of the source of dust relies on the surface winds exceeding the threshold erosion defined by the properties of the local surface. Most main sources of dust are dominated by depressions or inland drainage basins in arid regions because of the wind-erodible nature of their material surface as well as geomorphic dynamics (UNEP, WMO & UNCCD, 2016).

Indirect Drivers in Natural Ecosystems

Studies have stated that even though there is presently much ambiguity regarding the influence of human activity on dust storms, disturbance of natural systems as a result of human pressure is highly likely to increase in the coming years through human-induced climate change. Studies indicate that climate change impacts go beyond increased dust emissions. The advancement of roads, as well as lines of communication that divert or block drainage waters is another contributor to desiccation (UNEP, WMO & UNCCD, 2016).

Direct Drivers in the Ecosystems Dominated by Humans

Empirical studies have indicated that whilst a number of dust emissions come from natural sources (especially dry lake beds that have little human activity or vegetation), wind erosion is a crucial problem in human-dominated systems and facilitates dust storms. Studies have further indicated that the reduction of vegetation cover is a vital risk factor for dust storms and wind erosion (UNEP, WMO & UNCCD, 2016).

Impact of Dust Storms on Human Life and Green Areas

Dust storms have an impact in human life. Studies have stated that particle pollution such as dust impacts more people than any other pollutant because there is no safe threshold below which exposure to these mixtures of inorganic and organic substances and chemicals don’t threaten health (UNEP, WMO & UNCCD, 2016).

Cardio-Vascular Issues

Researchers have linked cardiovascular hospital mortality and deaths to windblown dust. This is also supported by numerous studies that have explained that emissions of mineral dust from arid lands contribute to numerous cardiovascular and respiratory illnesses. Studies have showed that dust is linked to an increased risk of heart disease caused by a limitation in blood supply to the tissues. In Japan, Asian dust clouds have been showed to increase visits to accident and emergency departments for cardio-vascular stress by 20 percent (UNEP, WMO & UNCCD, 2016).

Respiratory Disease as well as Asthma

Studies that were undertaken in Spain and Italy stated increased respiratory mortality amongst old people. Because asthma is among the leading non-communicable diseases in the world and affects more than 300 million individuals every year, special attention is being paid to the capacity of airborne mineral dust to exacerbate or cause asthma conditions (UNEP, WMO & UNCCD, 2016).

Eye and Skin Infections

Studies have stated that dust storms have an impact on skin and eye infections. It is noted that the contribution of Asian dust to skin and eye infections has been recognised to cause conjunctivitis, itchy skin and eyes. Even though the symptoms are not as serious, dust impacts nasal congestion and sore throats are also observed in otherwise healthy people (UNEP, WMO & UNCCD, 2016).

Meningococcal Meningitis

Studies have showed that meningococcal meningitis is spread by individual-to-individual contact via the respiratory droplets of infected individuals. Numerous investigations have showed that highly impacted regions demonstrate seasonal occurrences of cases that are strongly associated with levels of low humidity and greater concentrations of airborne dust during the dry season and outbreaks quickly diminish with the onset of the rainy season (UNEP, WMO & UNCCD, 2016). Even though the environmental connections of transmission are not completely understood, it is presumed that dust inhalation destroys the pharyngeal mucosa and triggers bacterial invasion. Studies have stated that epidemics may occur globally. On the other hand, the biggest outbreaks tend to happen in Africa. Studies have further indicated that the timing of the main outbreaks of meningitis and dust storms from the Sahara seem highly correlated. The UNEP, WMO & UNCCD (2016) hypothesised that the bacteria of Neisseria that are accountable for meningitis require iron-laden dust to develop and become virulent. Simulating and predicting windblown dust assists in understanding whatever role mineral dust might play in outbreaks of meningitis across Africa. Studies have showed that with improved understanding of dust's seasonality effect on meningitis, it is now possible to predict the seasonal outbreaks by estimating dust aerosols and providing information about the best way to target vaccination campaigns.

Valley Fever

On top of aspergillosis, airborne fungal diseases associated with dust involve valley fever. There are numerous endemic hotspots of valley fever in dry lands, especially southwestern United States, northeast Brazil and northern Mexico. Studies have stated that fungal spores in the soil related to valley fever are taken along with the desert dust, ready to be breathed in and infect people (UNEP, WMO & UNCCD, 2016). Numerous researchers have indicated that dust, presumably with cocci spores, might be carried over considerable distances. It is noted that considerable amounts of dust all over the border from Mexico to Texas during typical dust events in the area as well as dust sources in the United States impact the air quality of Mexico which is then inhaled by the people. Therefore, it is not known for how long the cocci spore might remain viable in the free atmosphere. A study conducted in California by the UNEP, WMO & UNCCD (2016) showed that blown spores in a storm infected individuals in Oregon.

On the other hand, dust storms also affect green areas. Studies have indicated that dust storms lead to reduced photosynthetic activity and the loss of plant tissue. This further leads to reduced reproduction within the green area. Studies have further indicated that dust storms lead to the delayed development of green plants. Increased soil erosion which results from dust storms triggers the process of land degradation which subsequently adversely affects green plants (UNEP, WMO & UNCCD, 2016; Tong et al., 2017).

Studies have showed that dust storms affect green areas because they cause a direct loss of tissue plant due to sand blasting by soil and sand particles. It is noted that because of the resulting loss of leaves, the plant experiences a reduction in photosynthesis and thus decreased energy (sugars) for the plant to use in reproduction, growing and the development of grain, fruit or fibre. In addition, the loss of plant energy for growth would also delay the development of the plant and in areas with short seasons of growing this could raise the risk of end-of-season drought by moving the moisture-sensitive period beyond the period of favourable rain and lead to lower yields in green plants. If there is sufficient soil and sand deposit material early in the season, young plants can be buried and possibly killed because of insufficient sunlight for photosynthesis (Sissakian et al., 2013). Studies have indicated that dust storms cause the loss of topsoil which increases soil erosion as well as further triggering the land degradation process and also desertification. This affects the growth of green areas because topsoil is the most fertile portion of the soil and contains numerous vital nutrients including phosphorous and potassium for the growth of plants and might be transported a long distance (Sissakian et al., 2013).

Dust Storms Come in Frequent Events

According to Wang et al. (2007), studies have stated that dust storms come in frequent events and historical experience could be used to predict future dust events. Researchers have indicated that meteorological variables such as low relative humidity, minimal cloud cover, maximum temperature, maximum surface winds, upper trough passage, frontal passage, maximum winds aloft as well as an upward vertical velocity usually occur as frequent events and, thus, have been applied to predict future dust events. It is further indicated that there are numerous variations as well as probable permutations of meteorological variables related to dust storm events such as the timing of different parameters. For example, an event might begin at night which might not result in any considerable dust blowing. Usually sufficient cloud cover can delay the start of a dust storm (Wang et al., 2007). Diverse combinations might only result in less intense events. It is noted that all of these possibilities must be taken into consideration and usually enable the prediction of future dust storms. Because of the dryness of desert air, there is a wide diurnal difference in temperature. According to Liu et al. (2010), empirical studies have showed that a rapid loss of heat at night from the trends of irradiative cooling lowers the inversion and settles the dust; consequently, dust storms generally lessen soon after sunset. When the formation of a surface-based inversion occurs, dust lifting is suppressed. On the other hand, for the dust that has already been suspended above the surface layer, a surface-based inversion will have less impact on its continued advection. Dust blowing from arid surroundings might be moved downwind over an extended region. Generally, predictions of visibility for dust storm events is very hard. Visibility might remain at 4-8km in dust mist and suspended dust for many days after the intense dust event (Wang et al., 2007). Powerful dust storms reduce visibility to near zero and in source regions visibility improves away from the source. It is noted that dust settles when winds drop below the speed required to carry the particles but some degree of dust haze persists almost constantly in the area during the dry season. Studies have indicated that the dust storms which occur as a result of bare soil and poor vegetation are associated with interference by individuals who have loosened the soil. When dust storms occur as a result of this process, it is very easy to predict future dust storms (Wang et al., 2007; Papayannis et al., 2008).

Studies have stated that dust events are more likely to be observed around major areas of desert. Studies have explained that wind erosion variability might potentially illustrate the environmental and climatological changes of the past (Wang et al., 2007). It is noted that periodic dust occurrences might impact ecosystem fertility and the spatial as well as temporal distribution of vegetation and animal species similar to climate change. Studies have further indicated that oceanic ecosystems have large amounts of nutrient-rich dust spread over vast regions where deserts are found close to the sea. It is noted that dust events in the Arctic (NE Iceland) are usually warm, occurring in the summer/autumn and during mild south-westerly winds, whereas the dust events of the subarctic (S Iceland) were usually cold and occur in the winter/spring when there are strong north-easterly winds (Wang et al., 2007).

Dust Storms in the Middle East

Dust storms refer to turbulent and a very strong winds that are capable of carrying clouds composed of fine dust, sand or soil and spreading these over large areas. A dust storm occurred in the Middle East in August 2015 that left seven nations in a thick cloud of sand and dust, obscuring them from being viewed by satellites (Hamidi, Kavianpour, & Shao, 2013; Jish Prakash et al., 2015). The dust storm that occurred was hazardous and hence it left many people injured, others dead and others with respiratory problems. Additionally, it caused several ports to close and airline flights were cancelled. At this time, the storm was linked to the civil war which was ongoing in Syria and the media spread this news to outlets in the United States, Middle East and Europe. Many reports about the dust storm claimed that it occurred due to land changes and cover; many farmers left their agricultural farms and there was an increase in military traffic. This is what was believed to have caused the large amount of dust that fuelled the storm (Hamidi et al., 2013). Other researchers concluded that the dust storm of August 2015 was caused by climatic changes and not human factors.

Researchers including Elie Bou-Zeid who is a civil and environmental engineer experienced this dust storm while in Lebanon and concluded that the unprecedented storm was not caused by human factors but rather by unusual weather as well as climatic changes (Najafi et al., 2014). The researchers say that uncovered land and soil disturbance caused the dust storm to take place. Bou-Zeid said that the link between the dust storm and fighting in

Syria was hypothetical, untested and no research at all suggested this was the case. There was no empirical data supporting the theory that fighting in Syria could cause the storm. As the dust storm continued receiving more attention, Bou-Zeid and Shmuel who belonged to the Agricultural Research Organization took the time to investigate the cause of the storm. They emailed several people to receive suggestions about this (Najafi et al., 2014).

The researchers published their study in November 2016 where they analysed data for vegetation cover for a number of months to investigate whether the military conflict in Syria had an impact on the land cover. From their observations, they found that the Normalised Difference Vegetation Index (NDVI) which refers to the measurement of how green a place is, was not found to be abnormally low (Alam, Trautmann, Blaschke & Subhan, 2014). Therefore, Bou Zeid concluded that there was nothing to prove that the place was more or less green during this season than normal. The dust storm struck the regions' vegetation land and it was very high as compared to 2007-2010. It was also more than 2001-2007 on an average basis. The data that were gathered for the wind speed, humidity and surface air temperature indicated that the climate resulted in the dust storm. Meteorological simulations which were run during and after the occurrence of the dust storm with the help of a Weather Research and Forecasting model (WRF) showed that the climate caused the storm (Alam et al., 2014). Bou-Zeid said that the simulations indicated that what was unique about this storm was the occurrence of the hot period and, therefore, the bare land with no vegetation was so dry and this made it easier to sweep away sand grains.

As compared to the last twenty years, the summer of 2015 was hot and dry in an unusual way. During August and September, temperatures were extremely high and there was low humidity. This was unusual compared to other droughts and vegetation level that had occurred between 2007 and 2010 according to Bou-Zeid. The amount of dust increased due to extremely arid conditions, lowering threshold erosion and causing a dislodgement of dust in the atmosphere. The storm was generated by unusual wind patterns which went from east to west instead of southwards from Syria and Iraq. This unusual wind pattern wind caused friction due to the reversal and hence the wind was directed westward before it was sent to the east Mediterranean coast.

In the Middle East, dust storms have increased significantly over the last fifteen years. Another reason for this increase could be Iraq where the flow of rivers has reduced due of dam constructions in the upstream countries (Gerivani, Lashkaripour, Ghafoori, & Jalali, 2011). The loss of marshes and the drying up of lakes in both Iran and Iraq leaves sediment behind, hence causing a huge amount of dust in the region. The situation regarding dust storms has been worsened by the presence of unsustainable mining and agricultural operations as well as oil construction. The UNEP has witnessed more than 300 dust events within a span of 10 years (Gerivani et al., 2011; Cao et al., 2015).

The occurrence of these storms has taken its toll in terms of health. The air becomes so polluted and hence when people inhale it, they are left with respiratory problems. This was reported by people who made calls as they coughed over the phone (Gerivani et al., 2011). People were forced to keep the doors and windows closed and make use of air-conditioners. Children were advised to play inside the house and not to go out because the wind could carry them away and injure them. In places such as Korea, China and Japan, many people developed health concerns which ultimately led to deaths. The storms which came from the Sahara Desert were believed to have spread lethal meningitis spores across the whole of central Africa. The dust carried in the wind caused lung illnesses once inhaled. It also caused heart diseases as per the report from the World Health Organization (WHO). WHO reports show that there were about seven million deaths attributable to dust storms each year. Pollution levels kept increasing up to 173 times, hence contributing to global dust of around 75% (Alam et al., 2014).

Method Used to Detect Dust Storms

There are two main methods used to detect dust storms: passive and active methods.

Passive Remote Sensing Techniques for Dust Detection

Visible as well as Infrared Method

Two separate methods were proposed by Chen et al. (2014) for the use of satellite-derived IR imagery and contrast reduction in visible imagery to recover desert aerosol optical thickness over land. Based on a desert model of aerosol and earth-atmosphere radiative transfer, Chen et al. developed a means of calculating aerosol optical depth over the sea utilising Meteosat visible data. They were also able to identify dust over the desert using a combination of visible and mid-IR solar channels.

Chen et al. (2014) proposed a guide for comparing dust storm intensity based on near-infrared wavelengths. They also proposed new ways of detecting dust over land and water using a combination of reflective and visible IR with a Moderate Resolution Imaging Spectrometer. In addition, they developed a new approach to distinguishing clouds and heavy aerosols using automated cloud algorithms classification. The system for detecting dust was based on reflectance analysis of visible channels or the brightness temperature in

IR channels. The magnitude of the difference in reflectance/brightness temperature is used to infer the dust particle signature.

Chen et al. also suggested that a tree decision classifier could be used to automatically detect dust storms by applying data from Moderate Resolution Imaging Spectrometer visible bands. This involves visual detection to remove datasets training of the tree decision classifier. Then hierarchical classifiers were recognised in relation to spectral features of the main reflecting objects. Then the spectral features of Moderate Resolution Imaging Spectrometer training datasets were analysed. Discrepancies in reflectance at 2.1 and 0.47μm was used to give the initial split level of the decision tree. The dust storm can be distinguished from other land-cover classes using band 7 (2.1 μm) (Chen et al., 2014).

Thermal Infrared Method

Chen et al. (2014) were able to identify soil-derived and volcanic aerosols by means of infrared observations at 12.0 and 11.0μm to enable dust detection at night and over bright surfaces. They suggested a means of detecting mineral dust according to particle size as given by TIR bands. The bi-spectral method of IR has been shown in empirical studies to be sensitive to loading dust but less sensitive to dust height. Chen et al. considered a different approach based on satellite observations.

Chen et al. (2014) suggested a new method to detect and extract dust storm information using multi-source remote sensing in the 12- and 11-μm spectrum. This approach produced meaningful outcomes for detecting and extracting data about dust storms but could not specify the intensity of the storm. Abuduwaili et al. (2010) devised a means of identifying dust outbreaks by examining differences in TIR temperature. However, the TIR method is greatly affected by the water vapor content and, therefore, should only be employed to reveal significant dust events when there is considerable water vapor content.

Chen et al. (2014) suggested a remote sensing technique for differentiating and detecting dust from cirrus by successive utilisation of the newly built up D-parameter in combination with the P-parameter. The P-parameter enables cirrus clouds to be distinguished from lower clouds, aerosol-filled pixels and clear clouds. Hao (2007) improved on the separation of airborne and cloud dust by developing a novel algorithm for over-water dust identification based on the aerosol identification method as used in the AVHRR imager applying 12.00-, 0.63-, 0.86- and 11.00-μm channels. Chen et al. (2014) established a dynamic reference BT difference algorithm to identify dust by removing the influence of temperature on the BTD.

Active Lidar Method of Remote Sensing

Studies have indicated that the active lidar-based technique might be divided into two parts: the scene classification algorithm and the selective iterated boundary locator algorithm (Chen et al., 2014). It is noted that the selective iterated boundary locator algorithm is adapted to identify a feature by use of a dynamic threshold scheme. The scene classification algorithm detects a feature as either aerosol or cloud based on two wavelength backscatter lidar profiles. Chen et al. (2014) derived an operational algorithm that is founded on multiple tests or on a single test which utilises a confidence function constructed from multiple-dimensional or one-dimensional PDFs to differentiate between aerosol and cloud.

On the other hand, studies have indicated that the mechanisms for detecting the outbreak of dust storms are not yet completely understood and, therefore, the methods for local dust storm prediction need further improvement. Studies have also stated that it is thus crucial to examine new techniques for predicting dust storms such as AI techniques. Studies have indicated that the application AI techniques has been recommended for predicting dust storms. An example of AI techniques includes artificial neural networks (ANN) (Chen et al., 2014).

Challenges and the Lack of Dust Storm Predictions in weather forecast

Studies stated that numerical prediction of dust using NWP-type methods faces numerous challenges. At the centre of the challenge are the enormous dimensions of scale needed to completely account for all of the physical processes associated with dust. The production of dust is a function of surface wind stress as well as soil conditions. Studies have indicated that wind alone may vary from generation of synoptic-scale to mesoscale phenomena including those generated by mountain passes or thunderstorms as well as micro-scale phenomena associated with boundary-layer mixing. In addition to meteorology, one should take into consideration the heterogeneity of the properties of soil and emissions physics. Normally, in global models the emission's functional form parameterisation is that of a power law in surface wind speed, causing the emissions to be highly sensitive to modelled wind fields. Consequently, size-dependent emissions and transport are significant factors of uncertainty. Studies have showed that the middling size of particles of dust that undergo long variety transport is surprising static, with an average diameter of 4-7 μm. The short-lived mode giant particles (15-100 μm) are a vital but woefully under-studied element of dust that contributes to the degraded quality of air and IR radiative impacts near source areas (Qian et al., 2013).

Studies have indicated that dust emissions' sensitivity to scale has resulted in recognition of the significance of model resolution. It should be noted that the modelled winds' quality is dictated by characteristics including horizontal resolution as well as numerical solver, and this quality is restricted by the comparatively low amount of wind observations available to compel the meteorological analysis triggering the simulation. Furthermore, almost all regional models and large-scale models don't have the ability to solve convective-scale phenomena and are thus possibly missing important emissions sources. In the past few years, a good level of accuracy for dust prediction at the synoptic scale and in a number of cases at the regional scale has been attained thanks to improvements in the model and in data assimilation to the point that the information might be delivered to the forecasters as guidance (Qian et al., 2013).

The expert reiterate that studies have indicated that dust storm prediction is a critical time task that needs to be finished in a restricted timeframe. For instance, a two-hour calculating limit is recommended for one-day predicting to make the outcomes relevant. It is noted that limited geographic region and/or resolution predicting is normally performed to complete the simulations within the time limit. On the other hand, a level of regional resolution is required for predicting dust storms to support emergency decision-making for government agencies including preparing medications for public health. It is noted that more than 12.7 hours is required to predict a 10 x 10 level domain size by use of a HPC cluster that has 25 calculating nodes. It is approximated that predicting the entire southwest United States with the domain size of 36 x 20 levels would require approximately 93 hours. It should be noted that such computing performance isn't acceptable because people would be predicting dust events that have already occurred (Qian et al., 2013).

According to Qian et al. (2013), numerous investigations have indicated that dust storms are disruptive events. Researchers have approximated that generally the total duration of dust storms in one year was less than 90 hours and also only accounted for less than one percent of one year assuming that each dust storms takes an average of two hours. Thus, a system of forecasting such events would expect diverse computing and access requirements at different times of the year and even diverse hours within a day (Qian et al., 2013).

Studies have stated that the availability of long records is crucial for improving the prediction of dust events. Any prediction and modelling requires enough data to verify and develop the model (Goudlie & Middleton, 2006). The historical data in the area are either in the media or not available, which makes them inaccessible. It is not surprising to find that most of the data in the hydrological services are kept on paper records. It is noted that such papers are easily damaged, leading to data being lost. It is also noted that a lot of historical data are also stored on computer tapes, making it difficult to read using the available computer facilities (Notaro et al., 2013). This poor state of historical data negatively affects the prediction of dust storms. Also, the majority of weather stations don't keep long records. The scarce historical data can't adequately provide historical patterns that might assist in predicting severe dust events (Goudlie & Middleton, 2006).

According to Qian et al. (2013), studies have further indicated that a crucial element in predicting dust storms is the human capacity for dynamical methods of prediction which is still very low. The dynamical methods of prediction are better suited to predicting dust storms because of their good performance in the short-term. On the other hand, these techniques require adequate information to capture the features of dust events. Studies have indicated that a forecast facilitates an early warning but might not give an actual "warning" if it is not correctly communicated. The failure to predict dust storms may lead to more severe impacts of these storms because people and policymakers are not warned about the dust storm and, therefore, might not be prepared for the consequences of dust storms. The inability to predict dust storms might not allow the National Centre of Meteorology and Seismology in the United Arab Emirates to give early warnings that update people regularly. Thus, people are unaware of the dust storms and this can result in severe effects. This website should update warnings every three hours giving information about dust conditions as well as visibility to the local media and the public. Therefore, due to the inability to predict dust storms, the website is unable to warn citizens to be prepared, aware or to take action regarding the severity of the conditions (Goudlie & Middleton, 2006).

Definition of Artificial Intelligence

Despite there not being a standard definition of the term "Artificial Intelligence," scholars and scientists have collectively defined it based on the way machines and software have been engineered to function. Black and Ertel (2011) approach the definition by separately identifying the broader meaning of "Artificial" and "Intelligence." Generally, they say that the meaning of the word "Artificial" is that it means it is not natural but man-made. "Intelligence" is the cleverness that machines display while carrying out duties. Therefore, AI is concerned with the development of machines (computers) and software (programs) that have the ability to perform tasks that would otherwise require the intelligence of a human being. These machines carry out tasks using intelligence indicators such as decision-making and task prioritisation.

In an attempt to present the definition even more clearly, Black and Ertel (2011) use a historical example given by John McCarthy to characterise the definition of AI. McCarthy's definition was that “the goal of AI is to develop a machine that behaves as if it were intelligent.” In testing the definition, a scenario of robotic vehicles manoeuvring through a limited space is given. These robotic vehicles move in such a manner that they avoid collisions while moving through the space and at the same time they follow a certain formation, moving carefully behind the “leader.” These vehicles are then described as being intelligent by McCarthy (Black & Ertel, 2011; see also Khalid & Jassim, 2014) .

Black and Ertel (2011) use psychologist Valentin Braitenberg’s description of such a complex behaviour, which can be produced by simple electrical circuits. Such circuits are used on Braitenberg vehicles which have two wheels, each driven by an independent electric motor. Light sensors on these vehicles interact in such a manner that they control the behaviour of the vehicles and the environment. These vehicles’ systems do not need any internal memory to operate, do not need any external interference, and are not affected by the environment. This is a representation of the simplest form of AI and the concept has been widely used in this field of study.

Dobrev (2004) looks at the definition of AI regarding the scope of computational models for human behaviour. According to Dobrev, the belief that human beings are intelligent underpins the rationale for making computational models of human behaviours; hence, such an indulgence results in the creation of AI. They say that one way of carrying out cognitive science is by doing experiments on humans so that how they behave in certain situations can be studied. This helps to establish if computers can be made that behave in a similar manner. Secondly, in order to achieve AI, computational models of human models can be developed. This is achievable using the strategy of affiliating with a person who carries out experiments that reveal what goes on inside people’s heads and then building computational models that mirror those kinds of processes (Dobrev, 2004).

Artificial Intelligence Techniques

Chen et al. (2008) identify 10 AI techniques. First, there is the case-based technique. This technique solves a problem by recalling similar past problems that have been solved by applying certain procedures (Lopex-Fernandez et al., 2011). It is assumed that the new problem has a similar solution. In order to successfully utilise case-based techniques, there is a need to have several past cases so that their solutions can be adapted in order to solve the new problem (Chen et al., 2008). Case-based reasoning involves four steps: retrieving the most relevant past cases, using them to provide solutions, revising the proposed solution, and retaining the solution for further use.

Secondly, there is the rule-based systems technique which solves problems based on expert knowledge. Conditions and actions are fed into the engine which uses a working memory of information regarding the problem. The system adds actions to the rule and by applying and using a match-select-act cycle, the working memory and knowledge base is repeated until no more relevant rules are found (Chen et al., 2008).

The third AI technique is genetic algorithms which are a search technique that mimics natural selection. Fourthly, there are cellular automata which are a series of dynamic models that are discreet in space, time and state. These consist of a regular lattice of cells which interacts with their neighbours according to local rules. The fuzzy system technique uses fuzzy sets to deal with imprecise and incomplete data to achieve AI. The rest of the techniques include multi-agent systems, swarm intelligence, reinforcement learning and hybrid systems (Chen et al., 2008). The suitability of a certain technique in environmental modelling is case-specific and these techniques have adaptable and diverse applications for the purposes of classifying, optimising and predicting AI. According to Chen et al. (2008), these techniques are suited to natural-resource management and exploring management strategies.

Hybrid Intelligent System

Hybrid intelligent systems are computational systems that combine at least two different intelligent technologies. Features of intelligent systems include adaptability, explanation ability, learning capability, imprecise tolerance, knowledge representation, discovery, and mining. These systems allow manipulation and symbolise different types of data and information that may originate from other sources as well as problem-solving and decision-making of various assignments. The main driving force behind hybrid intelligent systems is their strengths and weaknesses (Prentzas & Hatzilygeroudis, 2009). Therefore, incorporating the systems would lead to increased strength and fewer weaknesses. The most commonly known techniques that can be combined to form hybrid intelligent systems are the artificial neural network (ANN), expert system and case-based reasoning (CBR). There exist various types of hybrid intelligence systems that are parallel and cascade and are applicable in different scenarios.

Parallel Hybrid Intelligence System

A parallel hybrid intelligence system has two or more paths to pass or perform the required operation. All processes are done in a parallel manner. Hybrid intelligence systems are integrated to achieve the best results because some may be found to have explicit knowledge compared to others and, therefore, expertise automation. Other hybrid systems may be more experience driven, enabling them to match cases with the knowledge of domain vocabulary. When such systems are connected in parallel, they are ready or able to solve a number of problems. This type of hybrid intelligence system is desirable when every system is considered an expert in particular tasks (Panda, Abraham & Patra, 2012). Therefore, the systems are assigned tasks and they are expected to return outputs that are later integrated.

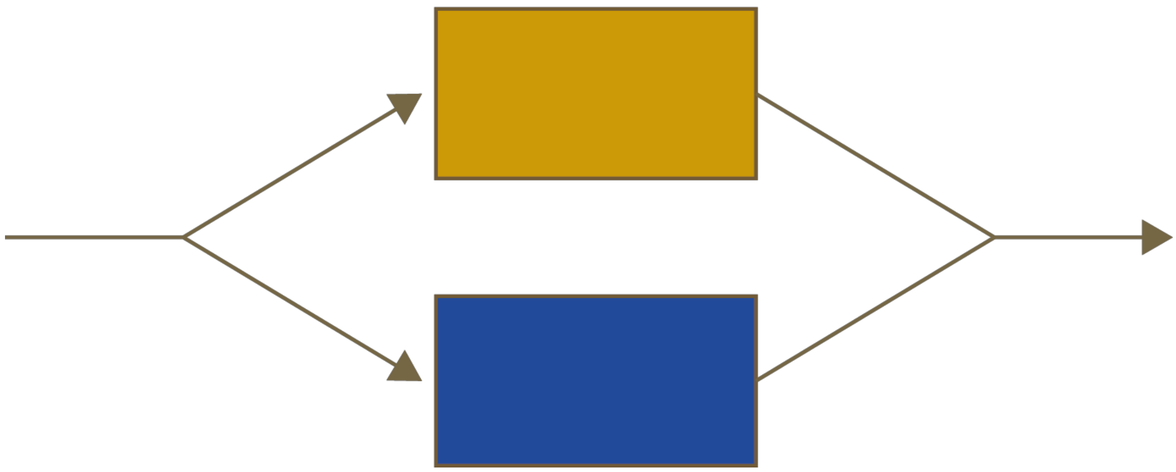


Figure 1. A sample of parallel hybrid intelligence systems (Panda, Abraham & Patra, 2012).

Advantages

There are several advantages with parallel hybrid intelligence systems. It is easier to add or remove a hybrid intelligence system from the connection, unlike in cascade integration where removal or additional of a new system can cause issues for the entire system. If one of the hybrid systems fails, the others continue functioning well because they do not depend on one another but are independent systems (Panda, Abraham & Patra, 2012). To design a parallel layout is very easy and, therefore, may not require a lot of time to be accomplished. This type of hybrid intelligence system is used when the problem that needs to be solved is large and is broken down into smaller portions that are handled separately. It uses divide and conquer algorithms.

Disadvantages

The disadvantages associated with parallel hybrid intelligence systems are as follows. First, the performance of one hybrid intelligence system cannot be increased because each is expected to work independently. This means the additional of new systems simply does not affect performance. The other disadvantage is that it requires a system integrator which is used to integrate the various responses from the systems to realise the full performance (Xiong, Zhang & Yin, 2009).

Cascade Hybrid Intelligence System

In a cascade hybrid intelligence system integration, the systems are connected sequentially in one path. All processes are done in a sequential manner. Many features are integrated to find solutions. They are able to retrieve cases together quickly as they narrow to search space. This type of hybrid intelligence system is desirable when there is a need for combined responses from all of the intelligence systems (Gemmis, Iaquinta, Lops, Musto, Narducci & Semeraro, 2009).

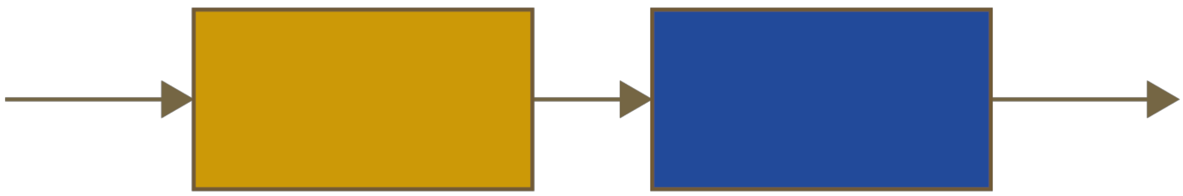


Figure 2. A sample of the cascade hybrid intelligence system (Panda, Abraham & Patra, 2012).

Advantages

Cascade hybrid intelligence systems have various advantages as follows.

There is increased performance of hybrid intelligence systems integrated sequentially. They are able to solve problems faster compared to hybrid intelligence systems that are integrated in a parallel manner. Adding new hybrid intelligence systems increases performance, although it is a somewhat difficult to add new intelligence systems in an existing hybrid.

Disadvantages

Cascade hybrid intelligence system integration has several disadvantages. When one hybrid intelligent system fails, all the other systems also fail to work because they are connected sequentially. The hybrid intelligence systems depend on one another, hence they all have to be in good condition to solve the existing problems. The other disadvantage of this type of hybrid intelligence system is the difficulties faced in adding or removing a system. The addition or removal of a hybrid intelligence system means that the whole architecture will be distorted and mended again for it to offer the same high performance as before (BORRAJO,Baruque, Corchado, Bajo, & Corchado, 2011). This type is not desirable for larger problems because it will slow down the whole process.

Case Based Reasoning

Researchers say that the field of CBR has a relatively young history which arose mainly for the purposes of cognitive learning. The focus of the technique was on how human beings could focus on a certain problem and generate a hypothesis based on present and past experiences and learning and acquiring new skills. This means that broadly construing CBR, it is a process through which human beings solve problems based on the solutions of previously encountered similar problems.

Some researchers give a chronological history of CBR in two geographical places: the United States of America and Europe. The history of CBR in the United States of America dates back to the 1970s at Yale University. This was when researcher and cognitive scientist Roger Schank first introduced it. In 1977, Shank created the first ever scripts for knowledge representation (Bergmann, 2000; see also Fielding, 1999). Further developments were made in 1983 when the first Dynamic Memory Theory and Memory Organisations Packets were implemented in a CBR system. Between 1983 and 1988, other systems such as JUDGE, SWALE and CHEF were developed at Yale University. In Austin, Texas, concept learning for CBR was developed by Bruce Porter. Between 1986 and 1989, he developed systems such as PROTOS which was an exemplar-based concept representation. Edwin Rissland of the University of Massachusetts worked on HYPO systems in collaboration with Ashley between 1990 and 1992, and on the CABARET system with Skalak in the same period. Jaime Carbonell and Manuela Veloso began working on prodigy analogy for Case Based Planning using analogy in 1990 (Bergmann, 2000; see also Fielding, 1999).

In Europe, CBR expert systems were developed by Michael M. Richter of the University of Kaiserslautern in Germany starting in 1988. Between 1988 and 1991, he developed the MOLTKE and PATDEX systems which were technical diagnosis systems. In

1991 and 1992, he developed the Case Based Planning systems CAPLAN and PARIS and European Projects INREA and INRECA-II. Ramon Mantaras of Spain worked on Case Based Planning learning for medical diagnosis in 1990. In Norway, Agnar Aamodt of the University of Trondheim worked on CBR and knowledge acquisition by developing the CREEK system in 1991, which is an integration of cases and general knowledge. Mark Keane of Trinity College, Dublin, developed the theory of analogical reasoning in 1988. Since 1991, there has been increasing interest across Europe in CBR research and several workshops have been organised for various research groups interested in the field (Bergmann, 2000).

Studies explained how CBR works using an example of a medical diagnosis. When a doctor is faced with a new patient, the doctor examines the patient's symptoms because this is the routine procedure. After doing this, the doctor compares the patient's symptoms with those of other patients who have shown similar symptoms. Because the doctor has treated other patients using certain medicines, it is only logical that the doctor will try to use the same medicine and procedures on the new patient. The same applies to other real-time situations, even in legal cases, such as determining the sentence of a felon by comparing his or her case with that of another whose case is almost the same.

Shiu and Pal (2004) say that most CBR systems' internal structures can be divided into two main parts: one that retrieves the case and the other that reasons out the possible solutions to the case. The purpose of the first part (the case retriever) is to search for similar cases in the system's database and select the most appropriate one. The second part's job is to use the case that has been selected to find a solution that will be applied to the presented problem. In this process, factors such as differences between the case being referred to and the one that has been presented are used to modify the retrieved solution so as to enable the system to reflect the differences. During the reasoning process, there may be further retrieval of other cases which may provide better solutions to the problem that has been presented.

CBR Methodology and Cycle

Experts say that the cycle of solving the problem in the CBR system has four main parts, all of which are applied categorically in order to arrive at the solution. The first part is retrieving, whereby similar cases that have been experienced in the past are located. Secondly, there is the re-using part, whereby cases are copied and their solutions are retrieved from the system. The solutions are then adapted to the new case in the revision stage. The final stage is the retaining of the new solution once it has been confirmed and validated by the system (Shiu & Pal, 2004).

Mascarenhas (2010) looks at CBR as a methodology that combines problem solving and learning. According to his explanation, this method solves new problems by remembering and adapting solutions to related problems that have been experienced in the past. He describes the technique as one which does not look for entirely new solutions but rather depends on “hints” from solutions that have already been implemented. A case in CBR is described as a “representation of an experienced situation which has its significant features, certain occurring contexts and the applied solution”. A problem, on the other hand, can be described as simply a new case that does not yet have a solution. According to the experts, CBR can sometimes be used as a synonym for analogical reasoning. However, the latter focuses on creating analogies across domains, while the former can be considered a form of intra-domain analogy (Wei & Tsatsoulis, 1993).

Studies presents the steps for conducting CBR in four steps. The first step, which is the retrieval step, aims to find the most similar cases from the database using a weighted sum of entities and returning all cases that fit within certain agreeable parameters. The re-use step involves adapting the retrieved solutions to the new problem using two methods: the transformational and derivational method. The transformational method involves modifying the former solutions using some transformational operators that are specific to a certain domain, whereas the derivational method involves reusing algorithms to generate the original solution so as to come up with a fresh one to fit the current problem (Mascarenhas 2010; Lopex Fernandez at al., 2011). The revision step involves evaluating the adapted solution and if it is unsuccessful it is repaired by applying knowledge that is domain-specific. Lastly, the solution is retained, and the new experience is stored as the new case by selecting the information to be stored and creating a new entry into the system’s memory for future reference to emerging similar cases.

CBR in prediction

Main, Dillon and Shiu (2001) say that CBR can be important in detection and prediction systems because it is a methodology that solves problems by utilising previous experiences, such as intrusions which may have occurred in a certain organisation’s system. Because the memory from past experiences is retained, it can be referenced for new intrusions which may be carried out in future. Once entry is identified as one which is not legal, the system uses the experience from the past intrusions to detect the current one. Main et al. (2001) give an example of a detection system for a ticketing machine. A CBR trouble ticketing system known as CRITTER manages and resolves network faults in a computer system. While the systems are being trouble ticketed, a ticket that is detected as a problem is kept until the problem is resolved by the system. The procedure for this is a description of the ticket's problem and provision of a solution.

Berkat (2011) elaborates how the CBR technique can be used in detection systems by detailing how the system can be applied to detect computer viruses. In this case, the memorisation phase is used to test if any problems – such as signatures of the already-identified viruses – are being recorded in the file that is being analysed by the system. If there is a detection of a part or whole of the signature, the file being tested is immediately considered to be infected. Secondly, the system can use adaptation, whereby it is assumed that a given file has been intruded by a virus which has a certain signature which had previously been removed by the system. In case there is a similar signature that is detected that looks like a copy of the already-removed file, the system signals that there is an intrusion. This method is known as the derivational adaptation technique, which derives the target solution from the source (Berkat, 2011). In computer virus detection, Berkat (2011) concludes that file analysis applies the same principles as the scanning method. He identifies the CBR technique as a tool that can be used to detect the viruses' signatures, storing them in the database and automatically updating the defence system by continuously detecting new threats.

Researchers describe the concepts, components and features of CBR and based on this, give the merits of using the same in detection. In assessing the advantages of using CBR as a useful approach in a number of situations, the applicability of CBR is supported by some of its characteristics. For instance, the cases that are to be solved normally occur frequently, meaning that the probability of getting a working solution is high. Again, the domain of CBR does not have to have an underlying model, and this flexibility favours its use in a number of cases (Shiu & Pal, 2004). There are also exceptions and novel cases that frequently occur, implying that IT specialists can profile cases by simply determining their signature, just like anti-virus databases do. The adaptation of past solutions to similar and relevant problems also provides a significant benefit in attacker and bot detection.

According to Shiu and Pal (2004), the use of CBR reduces the knowledge acquisition task through elimination of the need to extract the model designs or to follow a set of regulations. In CBR, the main idea is the collection of significant past problems and using them to solve new cases. Using CBR also helps the experts to avoid making mistakes similar to past ones all over again. CBR systems have the ability to record past mistakes and their probable causes, as well as the successful aspects of problem solving. If there is a repeat of the same mistakes that were recorded in the past among the new problems, the system can detect those and be able to prevent further damage. When using CBR for attacker and bot detection, there is flexibility in knowledge and modelling. Because various problems have their own unique formulas and models for solving them, CBR defines the scope of the problem. The domain experiences of CBR systems are the past experiences which are used categorically to provide reasonable solutions to existing problems. The system does this by adapting the solutions in an appropriate manner (Shiu & Pal, 2004).

In case the domain of a problem that is being solved has not been fully understood, or if enough information from similar past cases is not available, CBR systems can define and model the solutions. There are several situations that do not provide sufficient knowledge and this hinders the development of a working model, making it impossible to create a set of heuristics. CBR systems, in this case, allow for the creation of solutions by using the small amount of information that is available. This means that the user does not have to entirely understand a given case in order to be in a position to quantify the domain necessary for the new case. Additionally, CBR systems are able to make predictions regarding the probable success of the chosen solution. This is done by measuring the level of success of the solution in the past problem. When the solution is carried forward to the new case, the system is in a position to tell whether or not the solution will be successful (Shiu & Pal, 2004). One of the ways the system does this is by measuring the magnitude of the current query and weighing it against the past problem and also measuring the effectiveness of the solution in the past problem by using the stored information.

CBR also provides a means of explanation for the adaptation of a certain solution. The system can provide a number of reasons which convince the user and also justify the adaptation of a certain solution. This is because in the modern world of technology, users may have the urge to ask for the justification of a certain application and be assured of the quality of a given development before accepting it (Shiu & Pal, 2004; Otim, 2006). CBR does this by using similarities between the cases and the reasoning that has been used in the adaptation. At the same time, CBR can be used in more than one way in terms of prediction in different domain problem such as attacker and bot detection. This is because of the infinite ways that information can be represented, indexed, retrieved and adapted (Shiu & Pal, 2004; Kapetanakis et al., 2014). Another important advantage of using CBR is the fact that the systems reflect human reasoning in their application. CBR systems can approach situations in a similar manner that humans do, and the validity of the paradigm can easily be defended.

The communication between CBR systems and humans is validated, hence explanation and comprehension of various solutions that are adapted is possible (Shiu & Pal, 2004).

Uncertainty in Case Based Reasoning

Studies have indicated that uncertainty in CBR may happen for three main reasons. First, the information could be simply missing. For instance, the problem domain might be so complicated that it might only be represented partly. It is noted that even in simple domains it isn't always right to describe a complicated condition in each detail but to appear to apply the ease and functionality of the acquisition of the data represented in the case as a measure to determine the representation of a case. Secondly, for diverse problems, diverse characteristics and the challenge description will play diverse roles in attaining a solution. In other words, the significance of challenge description will not usually be the same. Thirdly, perfect forecasting is not possible. It is noted that people believe that there is no way to reduce or remove this kind of uncertainty from the challenge. The best that can be done is to choose a course of action in relation to what people understand and expect regarding the present situation and then keep track of the situation, learn more regarding the condition if possible, and regulate the actions dynamically (Wei & Tsatsoulis, 1993).

Researchers have proposed a novel methodological technique to CBR that enables it to apply decision theoretic techniques to handle multiple uncertainty types. Studies have indicated that the retrieval of old cases in CBR is regarded as a decision challenge, whereby each case from the case base gives an alternative prediction and solution for the probable outcomes for the present problem. It is noted that when uncertainty is encountered at the time of case-based problem solving, decision theory is used to assess each possible case in terms of the qualities that are important for the present problem so that the most appropriate old case might be chosen. Such incorporation gives a perfect accompaniment between CBR as well as decision analysis. It is noted that the set of cases or the retrieval of a case from the case base is regarded as a three-phase process. It is vital to note that at the time of the deterministic phase, the system detects the possible decision variables. At the time of the prognostic phase, the system detects the possible outcomes (Wei & Tsatsoulis, 1993). At the time of the ranking phase, a decision theoretic technique is applied to assess the best case(s) founded on the subjective probabilities allocated to the decision variables as well as the possible outcomes' utilities. If novel data become available, the system should determine if its previous decisions are consistent with the novel knowledge and, if not, perform the retrieval again. To show the use of decision theory in CBR there was the implementation of a CBR system that applies a decision theoretic technique to do case retrieval in cases of uncertainties. This system selects an initial set of cases founded on surface similarity, decides on the decision variables, builds up a decision model, evaluates likelihoods to decision variables as well as the utilities to outcomes, evaluates the decision challenge at hand, and lastly retrieves the best case. In the retrieval process, if any decision variable's value becomes known, the system will re-assess its decision to protect the most correct case chosen, thereby reacting to changing information regarding the world and the challenge.

Both CBR and decision theory have considerable limits; CBR cannot deal with uncertainty and decision theory is not good at optional generation and outcome forecasting. The work of Wei & Tsatsoulis (1993) lends support to the belief that an efficient consultation system which integrates methods of CBR and decision theory might be developed to assist CBR in handling uncertain conditions. This incorporation contributes in three ways: first of all, it makes the uncertainty challenge in CBR easier to cope with; secondly, it actively articulates, assesses, as well as evaluates a customised technique of the decision at hand which reflects the available alternatives of the decision maker, their genuine preferences, and their best information. Third, it improves the capability of a CBR system to solve challenges in which obvious consideration of trade-offs is important.

Studies have indicated that in order to solve the uncertainties in CBR, the integration of the CBR and decision-theoretic methods might be complementary. It is noted that decision theory assists CBR to deal with uncertainties in the domain problem, while CBR assists the decision theory to deal with complicated challenges with numerous variables. The aim of integrating CBR and decision theory is to enhance the ability of CBR systems to solve challenges in domains of the information that is not complete. The methodology perceives the retrieval of the old cases in CBR as a decision problem, whereby each case from the case base gives an alternative solution as well as a forecast of the probable outcomes for the problem. It is noted that when case-based problem-solving faces uncertainty, the methodology uses decision theory to assess each case in terms of the qualities that are important for the problem so that the most appropriate case can be chosen (Wei & Tsatsoulis, 1993).

The Future of Case Based Reasoning

Roth-Berghofer, Goker and Guvernir (2006) talk about the future combination of CBR and other methods. This will involve using the other methods as support systems, integration for quicker problem solving, and using CBR systems in a purely supporting role for the other systems. An example is case-based design systems which use supporting systems and reason maintenance systems. Additionally, another future development is improving adaptation capabilities, which has to be done in a completely autonomous manner. This will enable knowledge acquisition for the adaptation of rules. Moreover, future systems will be designed to scale up to larger problems. There have been efforts by designers to scale up CBR algorithms for application in research. The current technology is being designed to sufficiently integrate CBR in large-scale bases. Despite the fact that large scale bases are not mandatory for CBR, they would enable future systems to be applied in more than one area.

Roth-Berghofer, Goker and Guvernir (2006) looked at the future of CBR from the perspective of its use in robotics and automation. They focused their work on the retrieval and re-using steps in robotics systems, while working on concreting the use of CBR in robot soccer. The same is applicable in a number of general game-based adversarial environments. In the future, there will be a redefinition of the concepts of a “case” and the features that describe the states of cases by dividing them into two features: environment-based features and game-based features. In computing, the overall similarity and the similarities of the functions for various features are discussed. This is part of the continuing work on the extension of case description. The use of CBR for the achievement of self-healing in software systems, unlike alternative solutions, is one of the key future developments. The aim of developing future systems is to repair procedures that correct service failures that are triggered by external problems such as attackers and bots. Future systems may not require the availability of structured knowledge and new models of system behaviour will enable easier applicability and large-scale, complex systems. New tools that support the deployment and operation of self-healing infrastructures will enable CBR systems to be more efficient in future (Roth-Berghofer, Goker & Guvernir 2006). These developments will boost the applicability of CBR in additional real-world applications. CBR will be combined with several other resources and other reasoning paradigms which will enable integration with rule based and model-based systems.

Researchers provide a novel approach for studying customer preferences in content based CBR systems. Case-based learning is applied in combination with similarity measure learning. This could enables future system developers to be able to optimise the similarity measure directly, thereby improving the accuracy of the approximation of the unknown utility functions. This means that future systems will be able to solve new queries even if there is insufficient knowledge in the system memory to characterise the domain of the problems. This is because the systems will be able to improve the general quality of the absolute utility feedback which is required by the case-based learner to retrieve important information. This approach would be best suited for product recommendation systems that mainly deal with relatively static product databases. The risk of using wrong solutions to new cases will therefore be significantly minimised (Roth-Berghofer, Goker & Guvernir, 2006; Massie, Craw & Wiratunga, 2007).

Chance of Using AI Techniques to Predicting Dust Storms

Studies have explained that the mechanism for detecting the outbreak of dust storms is not yet completely understood and, therefore, the methods for local dust storm prediction need further improvement. Studies have further indicated that it is thus crucial to examine new techniques for predicting dust storms such as AI techniques. Studies have indicated that the application of AI techniques has been recommended for predicting dust storms. An example of AI techniques includes artificial neural networks (ANN). One of the main applications of ANNs is prediction. ANNs give an attractive alternative approach for both prediction researchers as well as practitioners. Numerous distinguishing features of the ANNs make them attractive and valuable for a forecasting task (Jamalizadeh et al., 2008).

Studies have stated that AI techniques such as the ANNs mimic the functioning of the human brain. It is noted that trials of modelling the biological neural systems' functioning may be traced back to the time of World War II. AI techniques such as the ANNs didn't become a viable technique until the early 1980s when the novel learning algorithms and also the good theoretical frameworks were developed along with faster computers which triggered a renaissance of interest in neural nets as well as their applications. Studies have stated that AI techniques now have a varied range of usage in meteorology such as the long term prediction of droughts and floods, short-term temperatures, events associated with El Niño, daily middling particulate matter concentrations, surface concentrations of ozone, and typhoon tracks (Jamalizadeh et al., 2008; see also Dagsson, Arnalds, & Olafsson, 2013).

Researcher applied a Back-Propagation model of neural network on a 0.5×0.5° net to forecast the large-scale occurrence of dust storms in China and achieved a satisfactory accuracy prediction. The formation of dust storm events is not usually accompanied by considerable large-scale changes in atmospheric circulation. The expert studied dust storms that were around the desert of Taklimakan from 1960 to 2000 by dividing the records of dust storm into 3 types: an ordinary dust storm process, a local dust storm process, and a regional strong dust storm process if dust is seen at 1, 2-7 or more than 7 stations, respectively. It is noted that according to the investigation by Jamalizadeh et al. (2008), local dust storm processes account for 6-11 percent of all dust storms and ordinary dust storm processes account for 41-61 percent. It is noted that because dust storms appear very regularly around the source regions of sand and most of the dust storms are due to the local flow conditions and thus can't be forecasted by large-scale coarse resolve meteorological models of dust storms, new methods need to be developed in these regions to forecast the occurrences of dust storms. The study have mentions four sites near rich sources of sand where selected and an ANNs software toolbox was applied which is one of the types of AI techniques to predict and simulate the occurrences of dust storms at these sites. The ANN model to be practically important for short-term prediction, they applied the previous day's local meteorological measurements as the independent variables and also the daily records of dust storms as the dependent variable. Studies have indicated that the advantage of the AI techniques such as the ANNs models over the models of multiple regression is that the AI techniques take into consideration the non-linear correlations between the independent and dependent variables between the environment control factors and dust storms.

Several researchers have shown better performances of ANNs in comparison to the models of multiple regressions. In this study they applied the variables chosen by the regression model as the inputs of the ANNs model and compare the imitation of the ANNs model with the regression model to examine if the meteorological variables have non-linear correlation with the occurrence of dust storms (Jamalizadeh et al., 2008).

Studies have indicated that some efforts have been made to improve forecast accuracy by using nonlinear and statistical models. It is noted that the traditional statistical methods of forecasting like linear multiple regressions are not able to capture the process of nonlinearity in climatic variables associated with dust storms. Numerous researchers have recommended that the forecasting strategy founded on AI is applied to forecast dust storms at a 1-day lead time. It is noted that with increasing populace density as well as a more mobile society, dust storms might pose a growing natural hazard (Jamalizadeh et al., 2008). Detection of their formation as well as their extension may be effective for their forecasting. It is noted that most meteorological events might influence visibility including local dusts as well as other aerosols, particularly in urban areas. The adoption of the expert system was undertaken which is able to learn the pattern of the events of dust storms so as to generate a detailed forecasting result. The study applied a multiplayer perceptron with back-propagation error algorithm to give predictions of dust storms. The ANNs technique was used for the hourly local dust event information for a specific region (Jamalizadeh et al., 2008).

It is noted that a feed-forward multilayer neural network comprises of an output layer and input layer. There are also more unseen layers in between the output and the input layer with some number of neurons on each. Because the intention of this study was to forecast dust storms, the initial condition is the rest of characteristics. It is noted that the network learns the correlation between the output-input data in the training set through network training whereby the weights are modified by presenting the training set's input-output patterns until a prescribed error criterion is fulfilled. It should be noted that these simple adoption experiments might give the practical results. In the forecasting of dust storms, the back propagation algorithm regulates weights of connection between neurons so as to reduce some of the squared errors between predictable outputs and perceived output values until an identified termination condition is met. It should be noted that this termination condition might be a threshold error or some loop count (epoch). Rate of learning and momentum constants are applied at the time of tuning weights in a step judgment. All of these constants and parameters should be modified to give the best result from the ANN. In fact, this implementation appears very easy in relation to the expert system. There was a need to predict all manifold features at the same time. This study indicated that for a selected hourly dust event observation, the relevant parameters are served to input neurons, whilst the next hour parameters are served to output neurons (Jamalizadeh et al., 2008).

Definition and Uses of Bayesian Networks

Bayesian Networks (BNs) are also referred to as Belief Networks and Bayesian Belief Networks (BBNs). The BNs are models of probabilistic graphics that represent a set of random variables as well as their conditional interdependencies through a cyclic directed graph. They might be applied to display and explore causal correlations between main factors and final results of the system in an understandable and straightforward manner. BNs are causal models they may also be applied to compute the effectiveness of the interventions including alternative management policies and decisions, and system changes including those forecasted for climate change. The uncertainties related to these causal correlations can also be examined at the same time. Studies have indicated that BNs are capable of maintaining clarity by making causal explicit assumptions and are usually applied for modelling when the correlations to be illustrated are not easily articulated by the use of mathematical notation (Ben-Gal et al., 2007).

The BNs belong to the probabilistic graphical models family. It is noted that these graphical structures are applied to represent knowledge regarding a domain of uncertainty. In particular, each node in the graph shows the random variable, whilst the edges amid the nodes show probabilistic dependencies amid the equivalent random variables. It is vital to note that these conditional dependencies in the graph are usually approximated by the use of known computational as well as statistical methods. Therefore, BNs combine the principles from probability theory, graph theory, statistics and computer science (Ding, 2010).

Studies have indicated that BNs appeared from an investigation into AI where they were initially developed as formal ways of analysing decision strategies under specific conditions. Studies have stated that the BNs are particularly important for different problems of variable size as well as complexity where suspicions are essential in the system. On the other hand, it is only lately that they have started to be implemented in the environmental modelling field. Numerous researchers have indicated that BNs use the Theorem of Bayes (also referred as Bayes' law or Bayes' rule) (Ben-Gal et al., 2007).

In Bayes' law, a prior (unconditional) probability reflects the probability that an input parameter will be in a specific state; the conditional probability computes the probability of the parameter's state given the input parameters' state impacting it. Also, the posterior probability is the probability that a parameter will be in a specific state provided the input constraints, the conditional probabilities, as well as the rules governing the way the likelihoods combine. It is noted that the linkage is resolved at the time nodes have been efficient by use of the Theorem of Bayes: $P(A|B) = P(B|A) P(A) P(B)$ (Equation 1) whereby $P(A)$ is the previous distribution of parameter A; $P(A|B)$ is the later distribution. Therefore, the probability of A provided new data B; as well as $P(B|A)$ the probability function, the probability of B provided existing data A. Studies have stated that BNs apply Bayes' Theorem to revise or update the beliefs of the likelihoods of system states taking specific values in light of novel evidence. Unlike many other techniques of modelling applied for environmental applications, BNs apply probabilistic instead of deterministic expressions to illustrate the correlations among variables. Deficiency of knowledge is taken into consideration in the network by the Bayesian probability theory's application. This enables subjective evaluations of the likelihood that a specific outcome will happen to be combined with more impartial data quantifying the occurrence's frequency in determining conditional probabilistic correlations. Because uncertainty is taken into consideration in the model itself, BNs are a specific suitable technique for handling the systems whereby uncertainty is inherent. That appears to be the main issue in ecological systems (Ben-Gal et al., 2007).

Studies have stated that uncertainty communication is also important when building up models for management. It is noted that BNs have some other appealing characteristics that make them particularly important for analysing data and decision-making. Furthermore, to their modest graphical causal structure they might be readily modified and extended; they might readily integrate missing data by the usage of Bayes' theorem; they are capable of being understood without much background knowledge of mathematics; they have been demonstrated to have good accuracy of prediction with small sample sizes; they might be applied to predict the possible values of states of the system provided opposing future scenarios; they might incorporate different sub-models, even if these work on diverse scales; and they might be easily united with decision tools of analysis to help management decision-making (Ben-Gal et al., 2007).

A study by Stephenson (2000) was undertaken to explore the application of BNs in forecasting rainfall. The expert developed a dynamic BN that represented the development over time of variables x and y; only three slices of time are indicated because Stephenson (2000) only considered the first order interactions amid the variables at diverse times (Stephenson, 2000). There are two diverse types of connections. Modern relationships are recognised amid the variables in the same time slice; Howes et al. (2010) assumed that all modern relationships are provided by the directed acyclic graph that is invariant over time. However, because the value of xt will usually be available from an ACM, the expert only considered non-modern relationships amid every rainfall variable ytk and the same variable at the following time step ytk+1. It is indicated that taking into consideration that for each instant t, that some values from xt and **y**t−1 are available as evidence at time t, the forecasting of rainfall was obtained by calculating the probabilities (Howes et al., 2010).

Probability in Weather Forecasting

According to Fraley et al. (2011), studies have stated that over the last two decades, the forecasting of weather has experienced a standard shift towards probabilistic predictions that take the probability distributions form over future weather events and quantities. It should be noted that the probabilistic forecasts enable optimum decision making for numerous purposes such as air traffic control, agriculture, ship routing, weather-risk finance and electricity generation. Until the early 1990s, a number of weather predictions were deterministic, meaning that just one "best" forecast was generated by a numerical model. It is noted that the recent start of ensemble forecasting systems marks a fundamental change (Fraley et al., 2011). An ensemble prediction comprises many numerical forecasts, each calculated in a diverse way. The statistical post-processing is therefore applied to convert the collaborative into sharp and calibrated probabilistic forecasts. Studies have showed that all forecasts undergo some level of ambiguity and with the application of ensemble forecasts and also statistical techniques that are increasingly developing the ability to approximate this uncertainty objectively (Fraley et al., 2011). It should be noted that for some prediction users it is enough to just take the standard prediction as the best probable approximation of what will happen but many users could possibly benefit more by understanding the vagueness and evaluating the risks. Among the best ways to express vagueness in a verifiable and consistent way is as probability forecasts. Studies have indicated that the probability forecast stipulates how probable the defined event is to happen as a percentage and might help users to evaluate the risks related to specific weather events (Fraley et al., 2011; Möller, Lenkoski, & Thorarinsdottir, 2013).

Forecast Evaluation

According to Root (2012), studies have showed that two aspects are vital in the assessment of weather forecasts which have numerical probability statements. The first aspect is the degree to which the predictions are distributed from the meteorological frequencies towards 0 and 100%. Another aspect is the degree to which the forecast probability agrees with the observed frequency percentage of an event's occurrence in the respective probability forecast category. Root (2012), in an assessment of subjective probability forecasts, referred to these two aspects as resolution and reliability respectively. It is noted that skills in both aspects are important to demonstrate skill in weather forecasting probability. The resolution of the forecasts might be assessed qualitatively by taking into consideration the association between the actual frequency distribution of forecasts probability and numerous other distributions. With impeccable resolution, the forecast probabilities would fall in the categories of 0 and 100% only. If in addition the predictions were all accurate, the forecast probabilities would be dispersed in these two classes in proportion to the observed rain or no rain cases (Root, 2012). With zero resolution, the probabilities forecast would fall in the class having the climatological frequency. A persistent forecast of the long-term climatology frequency of the rain would have agreed fairly well with the observed frequency in the period studies of 27%. On the other hand, the reliability of the forecasts might be assessed by noting the degree to which the forecast probability agrees with the frequency of rain occurrence in each category (Root, 2012).

Forecast Program

A study was conducted by Root (2012) regarding rain forecast probabilities. Starting with the winter season, the probability for rain expressed as a percentage has been regularly approximated for the routine Bay Area forecasts made at the time of the winter months. This study recorded the rain probabilities during the issuing of the forecast and was available on request for all forecasts issued in the winter season. It is noted that the rain forecast probabilities started as early as October in the season and lasted through to the month of May, changing with the level of the general winter rain period (Root, 2012). This study, on the other hand, considered only the predictions that were made during the months of January, February and December (the wettest months of the year). Generally, the forecasts were distributed four times per day and each forecast covered three 12-hour periods. This study indicated that the actual probabilities of rain were generally allocated 15-30 minutes prior to the filing of the forecasts. The rain probability allocated to each 12-hour period represented the best subjective approximation of the forecaster of the probability of rain occurring in the designated forecast period (Root, 2012).

A study by Root (2012) further developed two ways of approximating the ice formation probability on a roadway. It is noted that the simpler one assumes spatial dependence and the more complicated one models spatial dependence clearly. In this study, both probabilistic techniques significantly outperformed the raw prediction and the ensemble predictions the study considered and was almost half the total economic cost compared to relying on deterministic guidance. The two techniques give the same results because much of the spatial dependence in the formation of ice was already taken into consideration by the numerical weather forecasts. It should be noted that the method used by this study to the challenge of forecasting road ice goes beyond previous methods by giving probabilistic forecasts instead of deterministic predictions. In addition, it gives forecasts that are more reliable and calibrated as compared to probabilistic forecasts generated by ensembles of numerical forecasts (Root, 2012).

According to Root (2012), road ice is predicted by the use of mathematical models that generate physical interactions between the atmosphere and the road. It is noted that such models take into consideration meteorological parameters including air temperature, wind direction, precipitation, wind speed, dew point, humidity, and predict both road conditions and road surface temperature. On the other hand, in spite of the high degree of detail, their predictions are not usually accurate. Statistical techniques are used to select the weather and also other variables that best forecast road surface temperature. Likewise, statistical post processing techniques for numerical weather prediction outputs or road forecast models are proposed. This study by Root (2012) applied a back proliferation neural network to post process short-range predictions of road surface temperature. Root (2012) generated forecasts of road surface temperature through linear regression of the seen pavement temperature on the numerical predictions of the chosen climatological variables.

Current Application in Weather Probability Forecast using BN

According to de Kock et al. (2008), studies have indicated that probabilistic theory as well as statistics give a good theoretical base for handling uncertainty, a challenge typical of probability weather forecasts. On the other hand, past statistical methods have proven to be highly computationally complicated. The introduction of BNs gives a base to use statistical solutions in probability weather forecasts in a much more efficient manner (de Kock et al., 2008).

A study by de Kock et al. (2008) was undertaken to explore the application of weather probability forecasting using BNs. The aim of this study was to build up a dynamic BN to predict the weather. To evaluate how effective the generated BNs have proved to be, a sequence of experiments was undertaken to evaluate different aspects of their outcomes. Major factors to be evaluated are the forecasts' accuracy, the cost of computational learning and prediction generation and disparities between the generated BNs and how they are applied. It should be noted that in this study each experiment was undertaken on numerous diverse networks of differing size. It is noted that the number of nodes applied in each network was ten, twenty-five, fifty and one hundred. Studies have indicated that the networks also applied only the information for max temperature. Also, because of time constraints BNs could not be built up for all of the variables. The forecast capability must not change considerably between diverse variables so the study again gained insight into the possibility of BNs even with only one variable being applied (de Kock et al., 2008).

A study by Cofino et al. (2002) stated that the BNs' basic idea is to reproduce the most significant independencies and dependencies amid a set of variables in a graphical form which is easy to interpret and understand. This study took into consideration the subset of climatic stations whereby the variables for rainfall were represented pictorially by a set of nodes; one node for each variable. These nodes were linked by arrows that represent the cause and effect relationship. It should be noted that the directed graphs give a simple definition of independence founded on the existence or not of specific paths between the variables. The independency and dependency structure exhibited by an acyclic directed graph might also be articulated in terms of a joint probability distribution (JPD). Thus, the independencies from the graph are easily interpreted to the probabilistic model in a sound form. For instance, the JPD of the BNs defined by the graph requires the specification of 100 probability conditional tables, one of each variable conditioned to the set of its parents. Studies have stated that on top of the graph structure, the BNs require the specification of the conditional probability of each node offered its parents (Cofino et al., 2002). Studies have indicated that in the application of BNs in weather probability forecasting, once a model illustrating the relationships among the set of variables has been chosen, it might then be applied to answer questions when evidence becomes available. It is noted that before any data is known regarding the rainfall at the diverse stations, there is a priori or initial marginal probability for precipitation at each station k, P(yk = i), i = 0, 1, 2. These first probabilities might be efficiently computed taking advantage of the independence correlations encoded in the graph. Studies have stated that BNs have demonstrated their good accuracy for predicting with small sample sizes; they might be applied to predict the possible values of states of the system provided opposing future scenarios; they might incorporate different sub-models, even if these work on diverse scales; and they might be easily united with decision tools of analysis to help management decision-making when forecasting weather probabilities (Cofıno et al., 2002).

According to Gutiérrez Llorente et al. (2001), studies have indicated that models of BNs are the causal correlations between a set of variables. A BN has two main aspects. One of them is the graphical representation of the dependencies among variables. An application of an acyclic graph is used to represent this. Each variable is represented by one node within the graph. The representation of the direct causal dependencies is done with the help of the directed arc from the causing node to the affected node. Studies have showed that another element of the network is the gathering of the conditional probability tables that represent the probabilities of every state of a node provided the states its parents might take. It is noted that the strengths of correlations represented by the directed arcs might be demonstrated in the probability values kept within the conditional probability tables related with each node (Gutiérrez Llorente et al., 2001). It is further noted that these values are applied to infer the posterior probabilities of each variable provided those of its parents. Developing a model in this manner enables people to explicitly demonstrate the independencies and dependencies between the variables with a given domain. By using this structure as well as the Markov Blanket property's assumptions, all the dependencies are clearly modelled and people might restrict the statistical inference problem by using Bayes Theorem to only apply the applicable variables as evidence. Because of the structures enforced on the variables, the calculation needed for inference is greatly reduced from the case of performing statistical computation on the raw data. It is noted that the representation is efficient and compact, enabling statistical solutions to be used for much bigger problems than was previously possible (Gutiérrez Llorente et al., 2001). Because dependencies and variables are demonstrated in the graphical form, it is also easier to interpret, visualise and understand the relationships.

Integrated BN with CBR

According to Bruland et al. (2010), the main aim of incorporating CBR and the BN is that they both contribute to enhanced decision-making under uncertainty knowledge and incomplete information. CBR solves problems based on experience from other similar situations. CBR is founded on the model of human cognition and is meant to imitate our way of comprehending and predicting. It has obtained broad acceptance due to its psychological credibility and has been applied in numerous successful applications such as predicting the weather (Kiskac & Yardimci, 2004). Studies have further indicated that they are both supported as techniques that to some degree address problems in the weak theory domains. The weak theory domain is a domain whereby relationships between essential concepts are uncertain. General knowledge, with different levels of the theory, might usually be modelled by statistical distributions. The uncertainty type that deals with allocating the probability of a specific state provided a known distribution is known as aleatory uncertainty (Bruland et al., 2010).

Studies have explained that this type of uncertainty matches well with the BNs technique. Another uncertainty type, known as epistemic uncertainty, can be defined as the more general lack of information, whether weaker or stronger, and is connected to cognitive processing knowledge mechanisms. On the other hand, CBR may not the prefect choice to deal with aleatory uncertainty but is capable of using situation-specific experiences as among the epistemic knowledge type (Bruland et al., 2010). In this study, the lack of information and the unwell classification of the dust cases, was the reason behind applied BN approach first to solve these issues. Studies have indicated that the BNs institute a modelling framework made particularly for decision-making under aleatory uncertainty. Syntactically, the BN has a set of nodes, whereby each node represents a random variable in the domain and also where there are directed connections between pairs of variables. BNs might be applied for causal inference and for diagnostic inference. In CBR, a pool of past events or situations (i.e. the real episodes that have happened) make up the knowledge. The knowledge within the system of CBR is thus situation-specific, as opposed to the generalised knowledge in a Bayesian network (Bruland et al., 2010).

Studies have showed that the CBR system provides dust storm forecast text by using the repository of previous forecast information available in the case base. In this domain, a case comprises a pair of dust attribute values (for parameters such as humidity, the outlook and forecast time) and forecast information provided by human experts. CBR thus provides dust forecast information for new dust input data but might be easily extended for other dust parameters (Kiskac & Yardimci, 2004).

Retrieval

The information generation process in CBR starts with the retrieval of a case from the case base whose dust data is the most similar to the input. It should be noted that defining the similarity metric is thus essential for the retrieval element. The similarity calculation ensures that the retrieved weather data should have the similar number of states as the input. This is due to the fact that the number of states normally determines the number of phrases in the forecast information. The attributes of time are compared by using the differences between the time stamps in aligned dust states. Then there is the definition of similarity between the weather data mainly in terms of patterns across dust states. It is noted that the weather data and input data in each case in the case base are transformed into a representation demonstrating the pattern transition across dust states for each vectoral and scalar attribute (Kiskac & Yardimci, 2004; see also Al Murayziq, Kapetanakis & Petridis, 2017).

Reuse

The component of reuse of the CBR puts the prediction information related to the retrieval of similar dust data in the situation of the input. In order to do this, the prediction information is parsed to the detect attribute values from the retrieved dust data that are present in the information. It should be noted that these attribute values are then replaced with their equivalent from the input. The prediction information related to retrieved dust data might be reverted directly for output if the similarity amid the retrieved and input data equals to one at each step during the hierarchical calculation of similarity (Kiskac & Yardimci, 2004; see also Al Murayziq, Kapetanakis & Petridis, 2017).

Revision

The component of revision applies the expert rules to ensure that particular phrases conform to writing conventions in the domain. Those rules are learnt at the time of the post-edit task whereby experts are provided with input data and forecast information by the component of reuse (Kiskac & Yardimci, 2004; see also Al Murayziq, Kapetanakis & Petridis, 2017).

Retain

Studies have stated that retention might be undertaken in CBR whereby novel cases consisting of the output and input are added to the case base after the extra review by the experts. It is noted that the inputs whose prediction texts CBR was unable to generate might also be added after generation by the experts or by the use of other methods. CBR therefore evolves over the period of time and is capable of generating accurate prediction texts for most inputs when this element is functional. In order to combine the BN and CBR systems to solve uncertainty, there should be a split of the reasoning about the data problem according to their features. For instance, the BN is appropriate for global systems with uncertainty. On the other hand, the CBR is appropriate for local systems with numerous details and also a good likeness function, and for easily representing contexts to avoid exceptions (Kiskac & Yardimci, 2004; see also Al Murayziq, Kapetanakis & Petridis, 2017).

Studies have indicated that the incorporation of CBR and BN in solving uncertainties might be attained with either of the techniques as the slave and the other as the master, depending on which technique uses information given by the other. In this case, CBR is used as the slave. The recorded cases in the case base are utilised to compute the probability values related to each node of the BN. The effective inference algorithms exist for originating answers to questions provided the probability model expressed as a BN. Taking into consideration the relationships between features demonstrated by the network, these mechanisms of inference are applied to derive the most probable queries. Cases are applied to filter questions which were suggested by using Bayesian inference mechanisms (Houeland et al., 2010).

Studies have stated that CBR has been applied in identifying cases such as medical diagnoses. Expert has recommended a multi-stage structure for combining the fuzzy multi-criteria making of decisions and CBR with the objective of building a model to support intelligent decisions and carrying a use of tropical cyclone prediction that was proved to be successful. The recommended multi-stage structure is an extension of a task-based two-stage structure for CBR that applies the fuzzy multi-criteria decision-making technique to solve case selection uncertainty in CBR. It comprises three stages: case representation, case selection, and case adaptation and retention. In order to use the proposed framework for tropical cyclone prediction, Houeland et al. (2010) took into consideration a specific stage of Hurricane Alberto as a present case of a tropical cyclone. Provided the last observation of 24 hours of Alberto regarding its latitude, minimum central pressure, longitude, as well as maximum sustained winds, its position of track for the following 3 days might be forecasted by selecting the best analogue from the case bases from the previous tropical cyclones. In the stage of case representation, the cases are represented in terms of multiple criteria, multiple attributes as well as their order of importance (Houeland et al., 2010).

Kofod-Petersen et al. (2010) noted that for similarity evaluation, track positions are taken into consideration as the first most useful track position and attributes, and MSW and MCP as the second most useful set of attributes. The criteria for assessing the importance of previous cyclones are meteorological statistics, synoptic history and prediction critiques as well as satellite images. In the case representation stage, a subset of the historical tropical cyclones is chosen to depict the same past 24-hour track position (Kofod-Petersen et al., 2010). It is noted that the historical tropical cyclones have a slightly diverse structure as the present tropical cyclones might be regarded as either important or slightly important or very important to the present situation. The historical tropical cyclones which had been forecasted with huge errors might be very important to the present situation, possibly indicating the difficulty of predicting the future location as well as the intensity of the present cyclone. In the stage of case selection, these greater cases are considered accordingly and a weighted mean of attributes of the corresponding greater cases determines the intensity and position predictions for all lead times. The CBR technique for tropical cyclone forecasting has its advantages in the present stage if it may be further enhanced through improved knowledge acquisition, creation and reuse, capturing the knowledge of experts, learning from past experiences and conveying that knowledge for future decision-making (Kofod-Petersen et al., 2010).

The advantage of integrated BN with CBR

Studies have stated that the integration of BN and CBR offers numerous advantages. The benefit of integrating the BN and CBR is that the computations might be dynamically adapted. This means that people empirically identify when there is sufficient representative training information for the BN to generally generate better results as compared to the expert-modified CBR cases and might select the best-performing technique to resolve the following problem query.

Houeland et al. (2010) further reiterate that the integration of the BN and the case-based reasoning techniques could have more than one advantage: it can give extended functionality in addition to solving problems. Studies have also explained that the integration of CBR and the BN technique is mainly good at adding exceptions that might be applied to recover short-cut solutions rather than performing the normal reasoning. It is also noted that integrating the BN and CBR can answer numerous other questions regarding the statistical characteristics of the domain, including the probability of each possible grouping as a solution or the chance of a problem having specific attributes provided the observed outcome. Integrating BN and CBR makes it easier and simpler to solve the problem and also makes it possible to increase the accuracy of the classification (Houeland et al., 2010; Nikpour, Aamodt, & Skalle, 2017).

Definition and Uses of Rule Based System (RBS)

According to Ceccaroni (2000), the rule-based system (RBS) is also referred to as expert systems or production systems. The RBS is referred to as the model type of AI. The RBS applies the rules as the representation of knowledge for coded knowledge into the method. Numerous researchers have stated that the meanings of the RBS rely almost completely on the human expert's reasoning in solving knowledge exhaustive problems. Rather than showing knowledge in a static, declarative manner as a set of things that are true, the RBS characterises knowledge in relation to a set of rules which indicates what to conclude or what to do in diverse situations. The RBS is a manner of encoding the knowledge of human experts in a fairly narrow region into an automated system (Ceccaroni, 2000; Al Murayziq, Kapetanakis & Petridis, 2016).

The RBS is a form of applied AI that was built up by the AI community in the late-1960s (Alsaiari, 2012; Alves & Barboza, 2018). The fundamental idea behind the RBS is that expertise (which is an enormous body of task-specific knowledge) is transported from a human to the computer. It is noted that this knowledge is thereafter kept on the computer and those who are using it call upon the computer for particular advice as required. The computer might make inferences and arrive at a particular conclusion. Then, just like a human adviser, it offers advice and illustrates, if possible, the logic behind the advice (Alsaiari, 2012; Alves & Barboza, 2018).

Ceccaroni (2000) further reiterated that studies have indicated that the RBS comprises a set of rules of IF-THEN, a set of facts and a number of controllers governing the application of the rules, provided the facts. The expert system's idea is to apply knowledge from an expert system and to encode it into the set of rules. It is noted that when revealed to the same data, the expert system will work in the same way as an expert. Studies have showed that the RBS is a very simple model and might be implemented and used for a large variety of challenges.

The RBS's development started in the 1960s but became well-known in the 1970s and 1980s. One of the main interests in this area is the construction of the systems that could be founded on both the expert data and knowledge. Therefore, the construction methods may be divided into two main classes: data-based construction and knowledge-based construction. RBSs might be classified in the following elements: number of outputs and inputs, type of output and input values, the structure type, logic type, rule base type, number of machine learners as well as the computing environment type. For RBSs, both the outputs and inputs could be multiple or single. From this perception, RBSs might be divided into four main types in relation to the number of outputs and inputs: multiple-output-multiple-input, single-output-multiple-input, multiple-output-single-input, and single-output-single-input. All of the above four types might fit the features of association rules. This is due to the fact that the relation rules reflect the correlation between the attributes. A relation rule might have multiple or single rule terms in both the antecedent and consequential (right hand side) of the rule (Alsaiari, 2012).

Regarding the use of RBSs, studies have indicated that lately RBSs using AI have been applied to forecast the visibility, precipitation, marine fog, severe weather as well as other climatological conditions. RBSs on smaller scale models have numerous characteristic advantages that normally arise when AI techniques are used. First, they are taught and guided by human domain experts; thus, they become much more adaptive with this experiential guidance for applying knowledge acquisition methods (Alsaiari, 2012).

Rule Based System Methodology

Studies have stated that there are two main methods for the RBS: forward-chaining and backward-chaining (Cahn, 2014).

Forward-Chaining

The RBS, as its definition indicates, is adjustable to a variety of challenges. In some challenges, information is given with the rules and the AI follows them to see where they lead. Studies have stated that an example of this is a weather forecast whereby the problem is to get data based on a set of weather parameters (the working memory). Problems of this kind are solved by use of a forward-chaining, data-driven system that compares the data in the working memory over the states of the weather parameters (IF parts) of the rules and identifies the rules to fire (Cahn, 2014).

Backward-Chaining

Another method for the RBS is backward-chaining. In other challenges, an objective is stated and AI should look for a way to attain that specified objective. For instance, if there is a sign of a dust storm in a certain region, this AI could assume that a given region had the dust storm and try to regulate if its forecast is correct based on the available data. It is noted that this goal-driven backward-chaining system achieves this. In order do this, the system searches for the action in the clause THEN of the rules which fits the specified objective. This means that it searches for the rules that might generate this objective. It is vital to note that if a rule is initiated and fired, it takes the conditions of each rule as objectives and continues until either the available information fulfils all of the objectives or there are no more rules that match (Cahn, 2014).

Studies have indicated that of the two techniques available (backward- or forward-chaining). It should be noted that the evaluation of the conditions to activities in the rule base might help determine the chaining technique that is most appropriate (Veenendaal et al., 2015). If the ‘minimal’ rule has more situations as compared to deductions (that is the main hypothesis or objective can result in several further questions), forward-chaining is preferred. If the opposite holds true and therefore the middling rule has more deductions as compared to conditions in that every fact can fan out into a big number of novel actions or facts, backward-chaining is preferred. If none is prevailing, the actions number in the operational memory might help with the decision. If all actions are already well known and the aim of the system is to get to know where that data heads to, forward-chaining must be chosen. However, if no or few facts are acknowledged and the objective is to discover if one of several probable deductions is true, backward-chaining is preferred (Veenendaal et al., 2015).

Uses of rule-based system (RBS)

Studies have stated that the RBS has worked with other AI techniques such as Information Entropy Based Rule Generation to generate rules. Liu et al. (2014) divided the techniques of rule generation into conquer and divide and separate and conquer. Conquer and divide is a recursive technique because the rules generation is to choose an attribute to divide on and thereafter to recursively repeat the process for each branch covering the training set's subset. On the other hand, this technique has a principal disadvantage. Because the challenge emerges with the rule generation technique, the separate and conquer technique is encouraged to provide if-then rules iteratively and directly from the training examples.

Empirical research has indicated that the prism is a technique that follows the approach of separate and conquer. Liu et al. (2014) built up a modelled version of Prism known as PrismTCS. The inspiration is to increase computational efficiency because the initial Prism is computationally expensive. It is noted that the expensive computation results from the recurrent deletion of situations during rule generation as well as restoring training information to its original size for rule generation for the following class. PrismTCS usually selects the minority class as the target class. Therefore, PrismTCS induces the rules in the sequence of their usefulness without restoring the information to its initial size (in between the introduction of diverse rules). It should be noted that the prismTCS has been shown to generate grouping rules much faster but while maintaining the same level of projecting accuracy compared with the original PrismTCS (Liu et al., 2014).

On the other hand, researchers have recently indicated a number of limitations associated with the Prism algorithm in relation to the way the prism deals with clashes, underfitting of the notion in the training information and its computational efficiency. In relation to clashes, researchers have reported that Prism might generate some rules, each of them covering a clash set. A clash set comprises situations that belong to diverse classifications but can't be separated further. Studies have stated that according to the implementation of inducer software developed by Liu et al. (2014) for clash handling, Prism favours discarding a rule rather than allocating it to the majority class. It might lead to underfitting of the training set if a large number of rules is discarded. It is noted that for original Prism, this case might lead to a large number of situations remaining ungrouped because there is no evasion rule available and also the rules that cover the situations may be discarded. For PrismTCS, this case might make an evasion rule provide the wrong groupings to situations covered by the discarded rules. This is due to the fact that the rule of default is supposed to cover just the situations that belong to the popular class but inopportunely a number of rules covering other situations got discarded. In relation to computational efficiency, Prism favours discarding a rule if a clash happens. This shows that the algorithm consumes time to generate a rule that is finally discarded in a number of cases. It is equal to performing nothing and results in unnecessary costs of computation (Liu et al., 2014).

Researchers have recently developed another AI rule technique of generation known as ―Information Entropy Based Rule Generation that follows the separate as well as the conquer technique. On the other hand, it applies the from cause to impact technique while Prism applies the from impact to cause technique. The main focus of Information Entropy Based Rule Generation is on reducing the uncertainty which exists within the subset regardless of the target class. A common uncertainty measure is the information entropy that was introduced by Shannon. One of the benefits of Information Entropy Based Rule Generation compared with Prism can be observed from different examples with reference to the 24 dataset lens. Information Entropy Based Rule Generation captures information by conditional entropy. This is really unknown prior to the induction of the rule by the Prism algorithm. The PrismTCS would allocate class 1 as the target class to the initial rule being generated. Initial PrismTCS might also choose class 1 because the index of the class is smaller. However, the computational cost of the first rule generated by initial Prism is slightly higher than expected and, therefore, the rule has greater complexity. In a number of cases, the Prism algorithm might even be generating rules that are incomplete, covering a clash set, particularly if the target class is not a better fit to the attribute-value pairs in the training information (Liu et al., 2014).

Chapter Three

Methodology

Introduction:

A review of the empirical literature reveals the importance of AI and CBR and other techniques have been defined and described in detail. AI techniques have been applied in various fields. In this study it will be applied to the tasks of predicting future dust storms based on the history of the old dust cases. The technique will be used to predict new dust events using information and data recorded in weather stations and matching them by finding similarity with previous weather records. This tactic is based on the assumption that current or future weather data are similar to previously recorded cases and, therefore, the solutions will be similar. This means that the system will have a large database of previous dust cases which will be used to solve new dust cases. In general, BN will be applied to eliminate any uncertainty in CBR and then the four steps of CBR will be applied to the system for the prediction of dust. The steps will be the retrieval of the most relevant past cases, looking for a solution to the problems that were recorded, revising the proposed solution so as to ensure that the best one is chosen and then retaining the proposed solution in the system database so that it may be used in future. Finally, the RBS is applied to the BN-CBR prediction and executes the rules to generate the safety actions to reduce the harm of forthcoming dust events.

Rationale for Using a Bayesian Network (BN) with Case-Based Reasoning (CBR)

CBR is a logical choice for forecasting dust storms due to the features of the application area. CBR assumes that different problems which share similar features will have similar solutions . Consequently, CBR is well-suited approach for forecasting dust storms. Crucially, CBR not only forecasts events but also provides an indication of the probable success of a given recommendation (Kolodner, 1992). Again, this insight into the probable success of a venture is based on historical experience. For instance, when a solution is proposed, CBR is able to offer an indication of how successful that approach will prove to be (Pearl, 2000). In practical terms, this is achieved by gauging the scale of the current problem and setting this against historical experience gleaned from previous episodes while evaluating how effective the solution proved to be in the past.

One advantage associated with CBR is the flexibility that it offers owing to the ability to update the system with additional historical cases. These can then be used to enhance the decision-making process (Tran & Schönwälder, 2008; see also Bennacer et al., 2015). Moreover, a further advantage of CBR is its ability to accommodate incomplete data values and provide useful insight when faced with a large number of features. This is because CBR utilises informed guesses based on neighbouring values to fill in the gaps. The advantage of employing a BN when a model is thoroughly understood is that it is able to reason when faced with uncertainty. It does this without the need for a strong theory. Meanwhile, the key advantage of CBR is that it is able to reason when faced with uncertainty even when the model is not particularly well understood (Schiaffino & Amandi, 2000; see also Pavón, Laza & Luzón, (2009).

The modelling framework applied in BNs is well-suited for making decisions when faced with aleatory uncertainty. Each BN comprises a series of nodes. These nodes signify a random variable in the domain with links directed between pairs of variables. The combination of nodes and arcs give a certain acyclic graph structure. Details of the conditional independence statements in the domain are reflected in the links or the lack of links. These links can be regarded as giving details about the causal mechanism (Mascarenhas, 2010). BNs are increasingly being employed in order to explain causal inferences (Lacave & D´ıez, 2003). These BNs are well-suited to drawing out causal inferences and providing insight into diagnostic inference.

Within the confines of a monitored system, Bayesian reasoning is able to anticipate how likely certain events are to be realised. BNs not only utilise Bayesian reasoning but also illustrate the relationships between variables using graphical depictions. There are numerous examples of BNs in the empirical literature being used for purposes such as medical diagnosis and more generic modelling. For instance, Intellipath is a system deployed in the healthcare sector that exploits BNs for diagnosing conditions in pathology departments (Biocare Medical, 2009; Merli et al., 2016).

There are various types of network learning algorithm such as Gradient descent which is a first-order iterative optimisation algorithm for finding the minimum of a function and setting this against historical experience gleaned from previous episodes while evaluating how effective the solution proved to be in the past (Burges et al., 2005).

To find the local minimum of a function K2 is a well-known algorithm using a Greedy search to look for the structure with the highest score and it stops when no better structure can be found (Bouckaert, 1994). Hill climbing local is an iterative algorithm that starts with an arbitrary solution to a problem and then attempts to find a better solution by incrementally changing a single element of the solution (Tsamardinos et al., 2006). Tree augmented naïve Bayes (TAN) are probabilistic graphical models used for modelling large datasets involving lots of uncertainties among the various interdependent feature sets (Padmanaban, 2014; Терентьєв et al., 2016). Naïve Bayes are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features (Zhang, 2004). Markov blanket (MB) is the minimum set of variables assuming that the target is conditionally independent of the rest of the variables (Cheng & Greiner, 1999; Wong & Xiang, 1994).

In this study, three Bayesian network learning algorithms have been selected among others which include Hill climbing, TAN and Markov Blanket. This selection was based on the nature of the correction between the dust cases and the dataset used in this study. Each of the dust components are correlated with the others. The following sections provide more details about which suite of learning algorithms are appropriate for use with CBR.

The Path to Reach This Approach

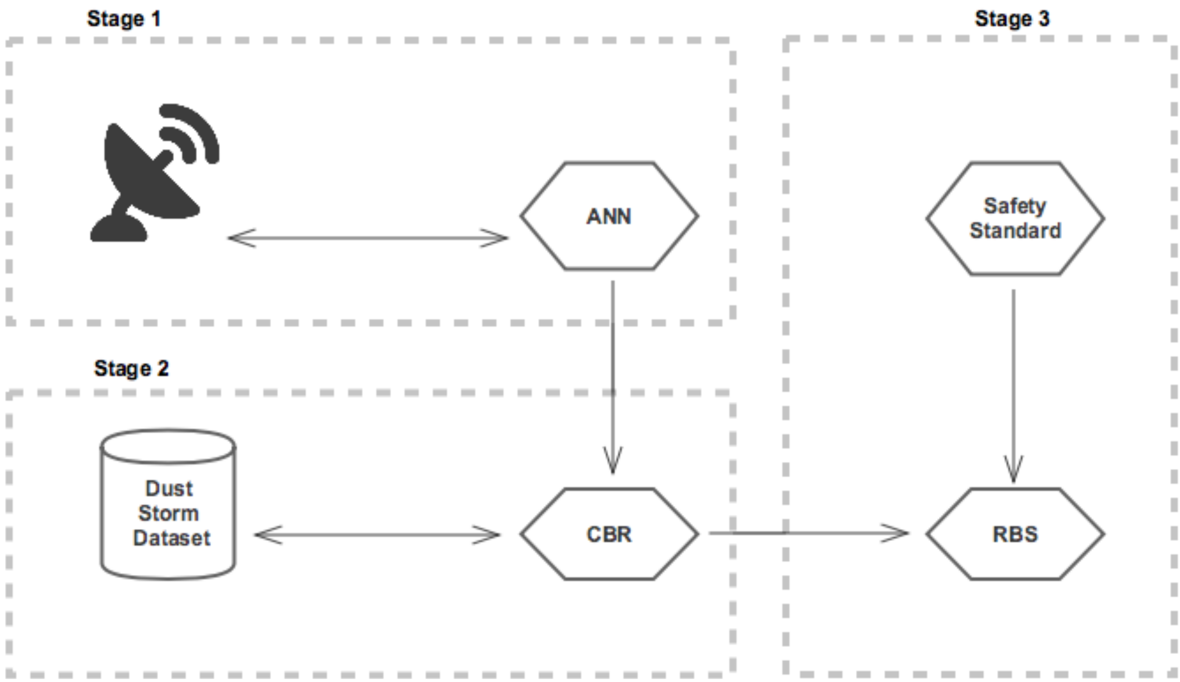


Figure 3. Initial proposed approach (Al Murayziq, Kapetanakis & Petridis, 2015).

Initially, this study planned to work on the identified main characteristics of dust cases using ANN and reclassify the old dust cases, using CBR to predict the coming dust events based on old methods. In addition, RBS was to be used to work in parallel to generate the rules and actions to take appropriate action to minimise the damage that would be caused by approaching dust storms. Having conducted the empirical literature review and examined the gap in the existing knowledge base, it is apparent that the ANN has been employed to forecast dust storms but in a different way. More specifically, the ANN has been utilised in order to confirm image resolutions, thereby making it possible to determine the size of dust storms on a map. In the initial test to proposed approach, the ANN is unable to reliably identify the key dust factors. Moreover, the ANN offers only a poor reclassification of cases considering the corrections of dust classes. In contrast, BN is based on probabilities and comparison of the results with those produced by the ANN in real-world settings met with the assumptions of meteorological experts. Given the nature of the project being undertaken as well as its aims, the CBR was selected from the available methodologies. Meanwhile, RBS was selected to develop a small expert system feeding predictions from BN-CBR results. The following section provides details of how these methods can be combined to develop a single approach to help forecast future dust events based on insight from the dust archive. The resulting model will provide insight into dust cases and the nature of the phenomenon. For instance, it will provide details about the frequency of dust cases, the regions that are particularly prone to experiencing dust storms, and the features that are most associated with the occurrence of dust storms. In addition, it will provide useful information that can be used to lessen the effects of dust events or even to help prevent them occurring at all. Moreover, the small expert system built by RBS will provide information regarding the most appropriate course of action to take when dust events occur.

Compare CBR with ANN (Artificial Neural Network)

An ANN is a computational model that is inspired by the natural neurons in the systems of living things. In animals, natural neurons receive signals by means of synapses that are located on dendrites or neuron membranes (Kriger, 2007). These signals are used to pass a signal and the networking of the neurons activates all of them once signals are passed continually. In computer applications, the basics are the inputs, which are similar to synapses. These are consequently multiplied by weights which is the strength of the received signal and lastly computed by a mathematical function which activates the neuron (Gershenson, 2014; Sharma, Rai & Dev, 2012).

In the recent past, methodologies have been proposed which apply the connectionist approach such as ANNs in designing CBR systems. However, while choosing which methodology is best for this study, the focus is placed on the outcome of construction which is dependent on a number of factors. CBR produces higher prediction rates than ANN. CBR is more flexible when the detection system is updated by adding the number of cases; it also produces better explanations for the choices (Priddy & Keller, 2005; Wilamowski, 2002). Additionally, ANN does not handle as many features as CBR; the latter can handle the missing data much better. However, when the system is well integrated and the two methods are combined, there can be greater accuracy and reliability. BN is involved with this approach to reclassify dust cases because some of the cases are not well classified. According to Bruland et al. (2010), studies have stated that the main aim of incorporating CBR and the BN is that they both contribute to enhanced decision-making under uncertainty knowledge and incomplete information (Bruland et al. 2010).

Compare CBR with RBS (Rule Based System)

CBR works without an explicit model. Cases that recognise the significant features are collected and concentrated to the case base during the development stage and after deployment. On the other hand, a RBS requires us to elicit an explicit structure of the domain but according to our experiences, acquiring knowledge is not a simple task and is associated with certain problems (Hüllermeier, 2007). The CBR model is much easier than explicit model creation; it creates a scenario whereby it is possible to create case bases without going through the problematic knowledge acquisition stage. An example of its application is found from domain experts who have acquired much knowledge due to experiences from the information technology field. When they are solving certain problems, it is extremely difficult for them to note the decision rules upon which they base their solutions but they find it easy to recall concrete cases that they faced earlier in practice. This shows that their state of mind is oriented toward the CBR model.

The CBR model is easier and much faster to develop compared to the rule based equivalent. Case bases work even when they are not fully completed; they are also easily customisable because non-information technology experts can also add cases to the structure. Case based systems are also straightforward and easy and their overall maintenance is simple compared to the RBS. Adding and removing rules from a RBS requires an entire system check for conflicting and redundant rules but this is not required with CBR because it does not require further checking and debugging (Hüllermeier, 2007; Richter & Weber, 2016).

Bayesian Network vs Lazy Learning in terms of classification

The dust storms were classified and categorised using lazy learning, to measure the classification accuracy compare to Bayesian network, but it is apparent that these results were less accurate than those obtained using the Bayesian classifier. With lazy learning, computation does not take place until after classification. In the training time, there is no attempt to devise theories such as decision trees. Moreover, predictions are arrived at by lazy learning algorithms using training examples (Webb, 1996).

Table 1: Summary of advantages of these three AI combinations.

An explanation for selecting BN-CBR with RBS	
Bayesian network (BN):	• Provides the probabilistic between different classes. • Capable of accommodating incomplete data. • Considers the nature of dependency between the attributes. • Helpful when choices need to be made and resolving any uncertainty in the data (Uusitalo, 2007).
Case based reasoning (CBR):	• Quickly arrives at possible solutions. • Easily explained. • Easy to acquire new cases. • Does not require complete data. • Lessons can be learnt both from success and failure (Haque et al., 2000).
Rule based system (RBS):	• The result is an expert system. • Arrives at workable rules. • Makes effective decisions. • Specific training is unnecessary. • Easily applied (McDermott, 1982).

Dataset Accessibility

This study has combined a BN with CBR in order to confirm whether improvements in forecasting dust storms can be realised by doing so. This method will be tested using past data for Riyadh in Saudi Arabia spanning the period 2008-2014 which has been provided by meteorologists in Saudi Arabia. It was in 1950 that the General Directorate of Meteorology and Environmental Protection was established in Saudi Arabia. In the intervening period it has served as the meteorological department responsible for the environment and climate change in the country.

In total, the dataset comprises 25 separate weather attributes, only 13 of which are required for the current study because these are the ones that are relevant to dust storms. Based on the findings, these 13 attributes have been found to be closely related to dust storms.

As previously stated, Saudi Arabia experiences large dust storms at regular intervals and these have adverse effects on industry, transport and the population in general. Furthermore, Saudi Arabia is particularly well-suited to studying forecasting ability in relation to dust storms because of the natural variation in its geography.

Dust case in CBR

After cleaning the dataset, the dust case contains these key attributes which are the key characteristics of the dust event.

S.N	Month	W.S.M	P	P.S	Rainfall	H	A.T	S.A

S.N Station number, which determines the region.

Month Month of the recorded dust case.

W.S.M Wind speed max.

P Pressure max/min.

Rainfall How many mm of rain in that day.

H Humidity max/min.

A.T Air temperature max/min.

S.A Surface ability.

P.S Pressure sea level max/min.

The key features of the dust has been ranked by a weather experts from the most important to the least low as follows:

P.MIN Pressure min.

W.S.M Wind speed max.

H.M Humidity max.

A.T Air temperature.

S.A Surface ability.

Rainfall How many mm of rain in that day.

Tools

MyCBR Tool for CBR

The available tools for developing the prediction system are free CBR and myCBR. The former is a free, open source Java implementation of a CBR engine, whereas the latter is a free-license. They both follow an approach whereby, for a certain attribute-value based case representation consisting of say n-attributes, the similarity between a query ‘q’ and a given case ‘c’ is calculated by the formula below:

$$Sim(q,c) = \sum_{i=1}^{n} \omega_i \cdot sim_i(q_i, c_i)$$

In the formula given above, sim1 and wi represent the local similarity measure and the weight of the attribute i . Sim denotes the global similarity measure.

For this study, MyCBR shall be used to reason out the cases and develop the structure for the system. The tool is a workbench that allows for the modelling and creation of highly sophisticated, knowledge-intensive resemblance procedures in a powerful graphic-user interface (GUI) (Bichindaritz & Montani, 2010; see also Bach & Althoff, 2012). It also allows for the integration of the information into the attacker and bot detection system by use of myCBR SDK (Software Development Kit). Additionally, the tool will enable the extraction of information and case based handling, besides extending to structured object-oriented case illustrations. The tool also has important taxonomy editors and fast prototyping through CSV. By using the myCBR workbench and the SDK, each attribute of the system can have several similarity measures.

MyCBR has three main features which are highly effective for application in this study. First, it has conceptual explanations which give information about the concepts of the particular domain of application. Secondly, it has forward-backward explanations which give a comprehensive explanation of the results of the retrieval process (Roth-Berghofer & Bahls, 2008 in Petridis & Wiratunga, 2008). It also has the feature of forward explanations which support the modelling of similarity measures. The tool has been developed as a rapid prototyping one with the broader purpose interface and also similarity-based retrieval engine. These functions enable quicker and easier integration with an application such as this one where the explanations can be adapted depending on the requirements.

Weka Tool for BN

The Weka tool was developed by the University of Waikato and is open source. This tool contains packages of machine learning algorithms for data mining. The algorithms can be used on datasets directly or can be imported from the user’s Java code. Weka tool can be used for clustering, classification, visualisation and viewing of datasets and shows the correlation between the attributes and regression (Hall et al., 2012). In this tool the BN algorithm has been applied by importing the dataset to the tool and applying the BN algorithm directly to the dataset to classify the dust storm cases and categorise them in different groups. This step eliminates the uncertainty in CBR.

IBM SPSS Modeller for RBS

SPSS Modeller is helpful for answering business questions by revealing trends as well as patterns in a structured manner to enable a deeper understanding of the constituents and customers (Modeler, 2011). SPSS Modeller provides flexibility and power in the interface of visual programming. It offers an entry point to the suite which is very competitive and affordable. The Modeller has functionalities such as the exportation and importation of data, preparation and graphical functionalities. SPSS Modeller reveals nuanced and complex patterns which are found in structured data. The Modeller provides algorithms that are advanced, automated modelling as well as data manipulation which helps to develop models that are predictive and are used to deliver business results in a quicker manner. It allows for wider connectivity of data which is linked to enterprise databases. It acts as a foundation of enterprise strength whereby there is predictive analytics which allows the Modeller server to be added and, hence, access openings to capabilities such as database mining and the SQL push back (Modeler, 2011).

SPSS Modeller has some powerful capabilities and techniques such as Text Analytics which are used in unstructured data sources such as RSS feeds, customer comments, documents and blogs which are in various multiple comments (Modeler, 2012). There are several benefits including the wide variety of advanced algorithms that contribute to better decision-making. It offers analytics of text and geospatial automated modelling. SPSS Modeller has gained predictive accuracy and it has been applied to hybrid and cloud deployment selections. The Modeller has the ability to conduct analysis without considering the data store; for example, flat file, warehouse, Hadoop, and database. The data is analysed regardless of the location, price, age, product or unstructured format; for example, emails, social and texts (Modeler, 2012).

Similarity Measurement Using the Euclidean Distance Algorithm

The underlying context of a particular application determines the meaning of similarity. There are, therefore, no fixed characteristics that can be applied to a certain comparative context. In this case, there are two main retrieval approaches. The first involves measuring similarity by computing the distance between the two cases and the second one deals with indexing structures of the case which is more suitable for text-based case applications. The Euclidean distance algorithm is based on the location of objects in the Euclidean space which is an ordered set of real numbers. According to Pal and Shiu (2004), the distance between the two cases is calculated as the root of the sum of the square of the arithmetical differences between two corresponding objects. The nearest arithmetical neighbours of a random case comprise the standard Euclidean distance.

The Euclidean distance algorithm takes two objects as an input and output as the measure of their similarity. The system utilises the Euclidean distance concept for similarity measuring. The data that are collected will be converted into feature data (using the features that were identified to be used in predicting new dust storms). Then a similarity measure method based on the Euclidean distance measuring concept is introduced with an accompanying attribute weight assigning method. The cases that are selected from the case base will be included in the case base of the CBR model and their profile documented. The effectiveness of each case is then determined by comparison by estimating the accuracy using a formulated principle devised by Pal and Shiu (2004). This is a concept that is based on the Euclidean distance measure method and involves searching the current input cases and then selecting the most appropriate referral case (dust case).

Approach of Research

Assuming that CBR can be used to forecast dust storms, in this study BNs are applied alongside CBR in a bid to improve forecast accuracy and enable members of the public to take steps to mitigate the adverse effects of these dust storms. The dust storms will be categorised and classified by means of the BN. In the event that patterns are identified in the data, these will be added to the CBR database. Future dust storms can be forecast by using CBR to match these storms with historical knowledge of such storms.

In order to improve on the current solutions, spatial data are required. CBR is particularly well-suited for these demands because it is able to explain case domain knowledge to experts in this field so they can arrive at the most appropriate solution to enhance the efficiency of their work.

The chosen methodology will enable meteorologists to recognise significant trends and appreciate the potential for the available solutions to detect dust storms in future. Any form of reliable early warning system would prove invaluable when helping to mitigate the effects of dust storms. For instance, prior knowledge of a storm would help to avoid financial loss in the farming community by taking appropriate evasive action. Analysing the available data will help the public authorities to devise systematic solutions in a timely manner.

It must be able to provide appropriate details relating to the weather in order to reliably predict the status. In This work, historical data was initially be applied in order to test that the chosen methodology is effective. Therefore, regions of the world that are regularly afflicted by dust storms will be selected for inclusion. For instance, the Middle East and China experience a relatively high number of dust storms and the frequency of these events has increased over time. For the purposes of this study, Saudi Arabia has been selected because it is a country with a coastline and also numerous inland cities, thereby enabling the research to establish what effect humidity has. In addition, details for the atmospheric conditions in Saudi Arabia are freely available and the data should, therefore, be relatively easy to collect.

More specifically, the research will focus on the coastal city of Damam and the inland city of Riyadh. After contacting the Saudi meteorological department, it has been possible to access weather records for both Damam and Riyadh for the period spanning 2008-2014. In addition to providing the various weather attributes, the records made available by the department include details of the processes used to clean the data. In total there are ten weather statuses that record around the clock. Each status represents a separate weather condition. For instance, W1=6, W2=5 shows that the first reading on that particular day indicated that it was windy and that it was subsequently hazy. These observations are made using satellite technology. The information gleaned from the Saudi meteorological department has been used to produce the following table which will prove useful when cleaning the data and describing the prevailing atmospheric conditions. Each code indicates a different form of dust condition.

Table 2: Dust categories description.

Type of Dust/Sand	Code
Haze	05
Widespread dust in suspension in the air, not raised by wind at or near the station at the time of observation	06
Dust or sand raised by wind at or near the station at the time of observation, but not well-developed dust whirl(s) or sand whirl(s), and no duststorm or sandstorm seen; or, in the case of ships, blowing spray at the station	07
Well-developed dust or sand whirl(s) seen at or near the station during the preceding hour or at the time of observation, but no dust storm or sandstorm	08
Dust storm or sandstorm within sight at the time of observation, or at the station during the preceding hour	09
Slight or moderate dust storm or sandstorm – has decreased during the preceding hour	30
Slight or moderate dust storm or sandstorm - no appreciable change during the preceding hour	31
Slight or moderate dust storm or sandstorm - has begun or has increased during the preceding hour	32
Severe dust storm or sandstorm - has decreased during the preceding hour	33
Severe dust storm or sandstorm - no appreciable change during the preceding hour	34
Severe dust storm or sandstorm - has begun or has increased during the preceding hour	35

With valuable input from experienced weather experts it has been possible to separate these weather codes into three separate categories. This will prove useful when cleaning and classifying the data. The codes 7, 8 and 9 indicate dusty conditions but no immediate suggestion that there will be a dust storm. Meanwhile, the codes from 9 to 32 indicate medium dust storms and it is advisable in these conditions to take appropriate steps to prepare. Finally, the codes 33, 34 and 35 indicate heavy dust storms and it is necessary to take action to minimise the adverse effects of such events. In order for the data to be assigned to the three categories, a short piece of code has been written in Microsoft Excel so that the various categories can be applied to the data. The code is presented below and works by checking each cell in the spreadsheet and assigning a category at the end.

```
Sub test()

For x = 2 To 7200

If (Cells(x, 14) = "7" Or Cells(x, 14) = "9") Or (Cells(x, 15) = "7" Or Cells(x, 15) = "9") Or (Cells(x, 16) = "7" Or Cells(x, 16) = "9") Or (Cells(x, 17) = "7" Or Cells(x, 17) = "9") Or (Cells(x, 18) = "7" Or Cells(x, 18) = "9") Or (Cells(x, 19) = "7" Or Cells(x, 19) = "9") Or (Cells(x, 20) = "7" Or Cells(x, 20) = "9") Or (Cells(x, 21) = "7" Or Cells(x, 21) = "9") Or (Cells(x, 22) = "7" Or Cells(x, 22) = "9") Or (Cells(x, 23) = "7" Or Cells(x, 23) = "9") Then Cells(x, 25) = "dusty" Else Cells(x, 25) = "No"

If (Cells(x, 14) = "30" Or Cells(x, 14) = "32") Or (Cells(x, 15) = "30" Or Cells(x, 15) = "32") Or (Cells(x, 16) = "30" Or Cells(x, 16) = "32") Or (Cells(x, 17) = "30" Or Cells(x, 17) = "32") Or (Cells(x, 18) = "30" Or Cells(x, 18) = "32") Or (Cells(x, 19) = "30" Or Cells(x, 19) = "32") Or (Cells(x, 20) = "30" Or Cells(x, 20) = "32") Or (Cells(x, 21) = "30" Or Cells(x, 21) = "32") Or (Cells(x, 22) = "30" Or Cells(x, 22) = "32") Or (Cells(x, 23) = "30" Or Cells(x, 23) = "32") Then Cells(x, 25) = "m-Dust"

If (Cells(x, 14) = "33" Or Cells(x, 14) = "35") Or (Cells(x, 15) = "33" Or Cells(x, 15) = "35") Or (Cells(x, 16) = "33" Or Cells(x, 16) = "35") Or (Cells(x, 17) = "33" Or Cells(x, 17) = "35") Or (Cells(x, 18) = "33" Or Cells(x, 18) = "35") Or (Cells(x, 19) = "33" Or Cells(x, 19) = "35") Or (Cells(x, 20) = "33" Or Cells(x, 20) = "35") Or (Cells(x, 21) = "33" Or Cells(x, 21) = "35") Or (Cells(x, 22) = "33" Or Cells(x, 22) = "35") Or (Cells(x, 23) = "33" Or Cells(x, 23) = "35") Then Cells(x, 25) = "h-Dust"

Next x

End Sub
```

Figure 4: Excel VB code to clean the dataset.

The proposed methodology for dust storm classification and prediction comprises three stages, In the final proposed approach, BN algorithms take the place of ANN, due the accuracy of classification, as is discussed in detail in the following chapters:

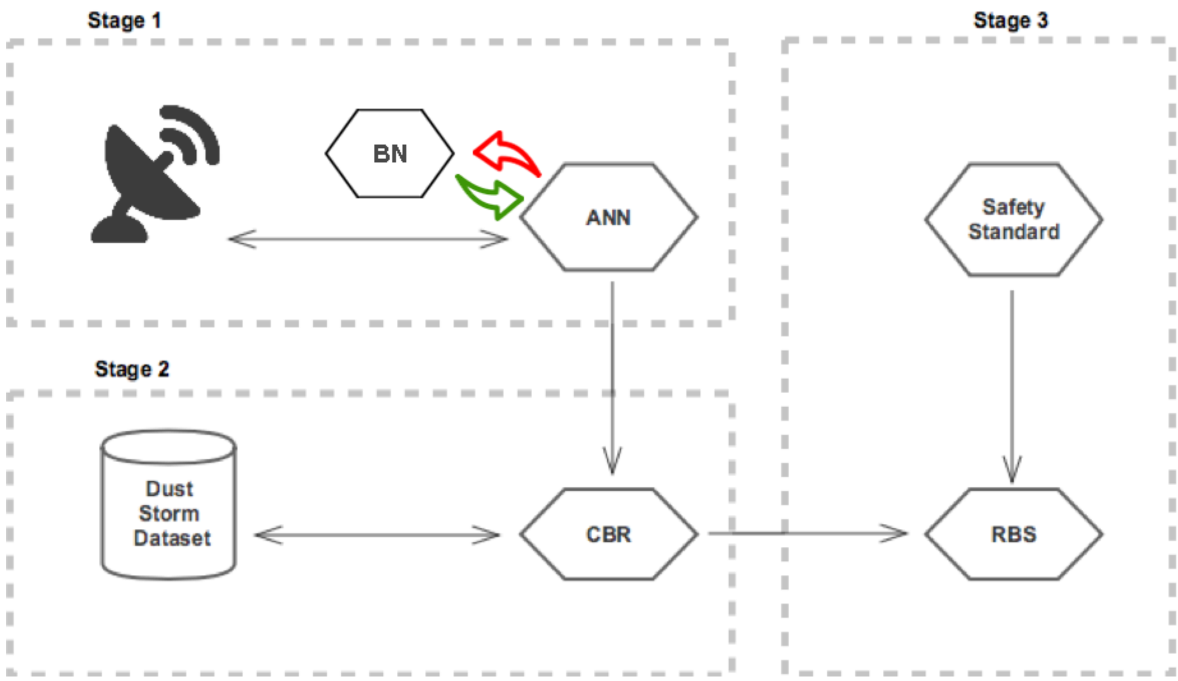


Figure 5. Proposed research methodology (Al Murayziq, Kapetanakis & Petridis, 2015).

Stage One: Get the weather forecast and use BN

For the purposes of the study, the term 'weather forecast' is not a prediction of what will occur in the following days. Rather, it refers to the current or expected values of each weather variable such as temperature, air pressure and humidity. The first process is to select the cases that will be studied (in this study two cities has been chosen, the capital city Riyadh and Damam city). Both cities experience regular dust storms, but their locations are quite different because Damam is coastal city on the Gulf sea and the other is inland. Therefore, this enables comparison of how dust attributes can be affected by the landscape. Specific stations have been chosen that are to the north of Riyadh and to the east of Damam. This is because the desert is located on these sides of the cities and, therefore, the dust must travel from these directions.

Secondly, after receiving the weather cases from the station, BN algorithm (1) will applied in Weka to these cases and old dust cases to classify them based on the probability to solve the uncertainty in the dataset and coming cases. For example, some cases have all the dust characteristics but in reality, they are not dust for logical reasons for example some cases are not well classify or the dust coming form another places. It will accurately classify these as not being dust cases to help with CBR prediction. Studies have stated that the BNs belong to the probabilistic graphical models family. These graphical structures are applied to represent knowledge regarding a domain of uncertainty. In particular, each node in the graph shows the random variable, whilst the edges amid the nodes show probabilistic dependencies amid the equivalent random variables. Deficiency of knowledge is taken into consideration in the network by the Bayesian probability theory's application. This enables subjective evaluations of the likelihood that a specific outcome will happen to be combined with more impartial data quantifying the occurrence's frequency in determining conditional probabilistic correlations. Because uncertainty is taken into consideration in the model itself, the BNs are a suitable technique for handling the systems whereby uncertainty is inherent and appears to be the main issue in ecological systems (Ben-Gal et al., 2007).

$$P\ (E|F) = \frac{P(F|E)\ P(E)}{P(F)} \tag{1}$$

Where P(E) ≠ 0 and P(F) ≠ 0

The BN process employs historical dust cases that will be classified according to the new weather data. By doing so, the new weather records will be assigned to a dust category, thereby enabling an initial forecast to be made. The forecast can be measured using the precision and recall measures. The prediction will then be tested using the CBR.

BN is primarily employed to reclassify the data. This is necessary because the dataset comprises many instances in which the weather status is not recorded. The following dust attributes are required for the BN: minimum air temperature, maximum air temperature, wind speed, minimum pressure, maximum pressure, minimum humidity and maximum humidity. The BN algorithm combines these attributes as inputs that are used to classify the dust events. The aim is to produce a weather status for the day. For instance, for each case the algorithm will classify a weather status by estimating the likelihood of a dust event in any of the categories. As such, the BN will determine whether or not the cases have been assigned to the appropriate category. If necessary, cases will need to be transferred from one category to another if their weather status has been inaccurately recorded.

Different BN structures have applied and tested on the dataset, in order to choose the right BN node structure that fit with dust storm characteristic, K2, Markov and TAN are take part in this study to build up more effective and reliable predication. The BN algorithm was chosen because it is more accurate at classifying the data into the correct dust type based on the confusion matrix. In addition, precision and recall measurements are more accurate when using the BN compared to the alternatives. For instance, the multilayer proportion which is part of the ANN gave good classification results but the majority of cases that were classified accurately were in just a single category with more cases being added to other categories. Therefore, the decision was taken not to proceed with the ANN.

The results of BN are compared to lazy learning. The confusion matrix revealed that the lazy learning algorithm was less accurate than the BN at predicting heavy dust events and it also had a higher mean absolute error than the BN. The accuracy of the Kstar algorithm also underperforms the BN in terms of precision and recall. Once the BN has been applied, the reclassified data must be exported from Weka tools and saved as an Excel file. Once it is in an Excel format it can be used in the CBR cycle using MyCBR tools to help forecast future dust storms.

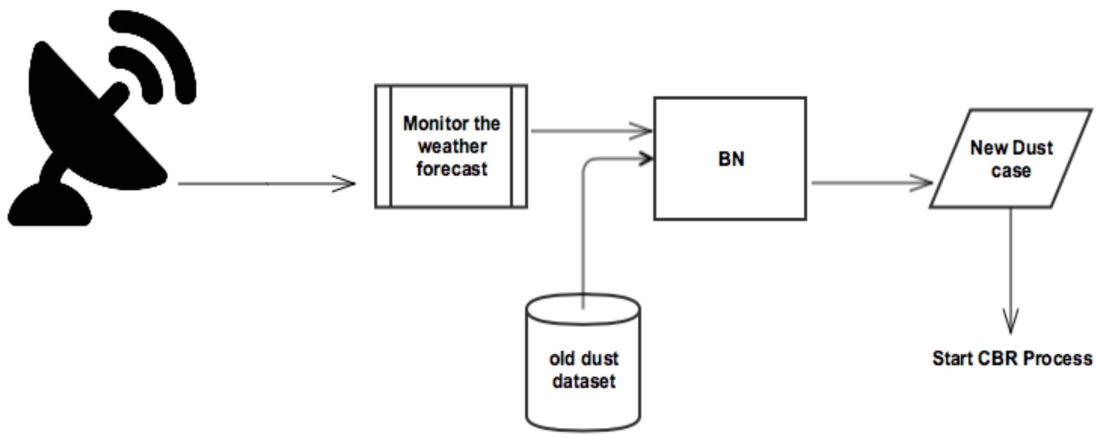


Figure 6. Applying BN in stage one (Al Murayziq, Kapetanakis & Petridis, 2015).

Stage Two: Identify the Dust Attributes and Apply CBR

Stage two of our methodology establishes the current benchmark of any existing cases for better understanding the current dataset of dust storm cases. The dust case contains the following attributes which are the key characteristics of the dust event and the case content one layer. The dataset was collected by daily monitoring from Riyadh radar station with register 403345 and it contains more than 7000 dust cases, while fewer than 6000 cases are collected from Damam at radar station register 504453.

S.N	Month	W.S.M	P	P.S	Rainfall	H	A.T	S

- **S.N** = Station number**.**
- **Month** = Month.
- **W.S.M** = Wind speed max km/h**.**
- **P** = Pressure in mm Hg.
- **P.S** = Pressure sea level in mm Hg**.**
- **Rainfall** = Rainfall in mm.
- **H** = Humidity max and min in H.
- **A.T** = Air temperature max and min in C.
- **S** = Surface ability.

In order for the CBR to be applied successfully, it is necessary to identify the attributes that have the most significant influence on the problem domain. There are two main ways in which the most significant features of dust events can be identified. The first of these is to rely on advice offered by weather experts at the Meteorology department in Saudi Arabia. They use the following ranking system to explain the influence of the various weather phenomena on dust events. The different features are ranked in terms of their importance from high to low:

- **W.S.M** = Wind speed max.
- **P.MIN**= Pressure min.
- **H.M**= Humidity min.
- **A.T**= Air temperature max.
- **S** = Surface ability.
- **Rainfall** = No rain.

The second approach is to apply the attribute ranker which is a Weka tool. The attribute ranker utilises the information gain algorithm (2) to estimate the information gain relative to the stated class. The results are broadly similar to the ranking produced by the meteorological experts but a number of the attributes are in slightly different ranking positions according to the location of the city which would influence factors such as sea pressure and humidity. However, these modest differences will not adversely influence the similarity matrix to any significant extent.

$$\text{InfoGain(Class,Attribute)} = \text{H(Class)} - \text{H(Class} \mid \text{Attribute)}. \qquad (2)$$

After the BN algorithm has been run and the dust characteristics identified, CBR will take place to apply this cycle on the dataset using the MyCBR tool starting with the retrieval of a case from the case base whose dust data is the most similar to the input/target. It should be noted that defining the similarity metric is thus essential for the retrieval element. The similarity calculation makes sure that the retrieved weather data should have a similar number of states as the input. In the retrieval step the CBR will apply the weighted Euclidean distance to find out the most similar case to the target by ordering the whole dataset from the closest cases to the farthest cases. The Euclidean distance algorithm (3) is based on the location of objects in the Euclidean space, which is an ordered set of real numbers. According to Pal and Shiu, the distance between the two cases is calculated as the root of the sum of the square of the arithmetical differences between two corresponding objects (Shiu & Pal, 2004).

This stage involves using case-based reasoning for the forecasting coming dust events regarding the retrieval of similar dust data in the situation of the input. This is achieved by parsing the forecasted information in order to identify attribute values from the dust data that has been retrieved from the information. In effect, insight into new cases is being provided by observations of historical events. This fits with the underlying hypotheses of CBR that similar problems have similar solutions. There may well be more than one solution in the reuse stage. It is for the client to assess these solutions and determine whether they need to be altered in any way before they are used in future. It is then necessary to revise what has been done to ensure that all phrases are suitable for the writing conventions in the particular domain. For instance, with regards to dust events, lessons gleaned from historical dust events can be used to reduce the negative effects resulting from future dust storms. Furthermore, the solution should be retained with reference to the case base. The empirical literature suggests that retention can be achieved using CBR whereby the outputs are novel cases. Inputs are added to the case base following review by trained personnel to ensure that the available solutions have been fully evaluated and are appropriate for the different categories of dust events. One of the benefits of using the retaining stage in the CBR cycle is that it saves all of the various solutions for dust events. Therefore, this makes it easier to learn from past examples and evaluate what actions need to be taken in future. Furthermore, the retaining stage could provide useful information for future research.

Once MyCBR has been completed, the CBR process needs to be applied to the BN dataset. Once the degree of similarity has been established and the dataset has been ranked in terms of size, the nearest neighbour 3NN is applied. KNN helps to arrive at results in terms of the solution that most closely matches the incoming dust events. For instance, in target cases the weather status is not known and, therefore, the odd KNN number is used to vote on the closest cases. It is the voting system that is the main benefit of KNN because this confirms the weather status/target. The accuracy of CBR forecasts using 10-folds refers to the fact that upon receiving the result of the weather status, the step is applied 10 times in 3NN. This involves choosing a number of cases at random and not immediately revealing their weather status. As such, the accuracy of CBR forecasts is tested by observing how well it is able to correctly anticipate the weather status. The number of correct forecasts relative to the number of inaccurate forecasts gives a measure of the CBR's efficiency at forecasting future dust events. This process is repeated several times but always using an odd fold number to check whether the efficiency score alters and to confirm how accurately the CBR can forecast dust events.

The solution to the similar cases that matched the forecast case/target will reuse the new dust cases so as to minimise the harm that dust storms cause to both property and health. However, one problem that is apparent is that the dataset obtained from the meteorology department for Damam and Riyadh does not include the solutions. Therefore, in order to overcome this issue, a rule based system (RBS) is required to establish the rules that are most closely associated with dust events. These must then be matched with the action to provide the solution. The actions are derived from the weather phenomena safety standard as applies in Saudi Arabia. These actions are evaluated by weather experts so they can be applied in real-world settings. It is because of this step and the CBR cycle that the hybrid intelligence system has been applied in this study.

$$D(x,y) = \sqrt{\sum_{i=0}^{n} wi(xi - yi)^2} \tag{3}$$

Failure Dust Storm Cases

Logically there are some dust cases that do not match with our expectations. This could be due to some incorrectly recorded data or the dust being carried from other places. Failure cases mean the dust cases fail to match the final result of our prediction. Due to the nature of the problem domain, there is no chance to train the BN to re-correct the failure cases because dust storms that are not matched correctly and predicted well could lead to false predictions and false warnings. In our datasets the failure cases are rare and in the CBR experiment we eliminate these failure cases by applying the BN classification algorithm. This step gives an advantage to limit the experiment in testing the dust cases that occur from one place and evaluate our prediction accuracy. The CBR process will deal with most of the real weather recorded. That means that all of the predictions will be based on the input. As long as they are clean and identify well, the prediction will be more realistic and effective.

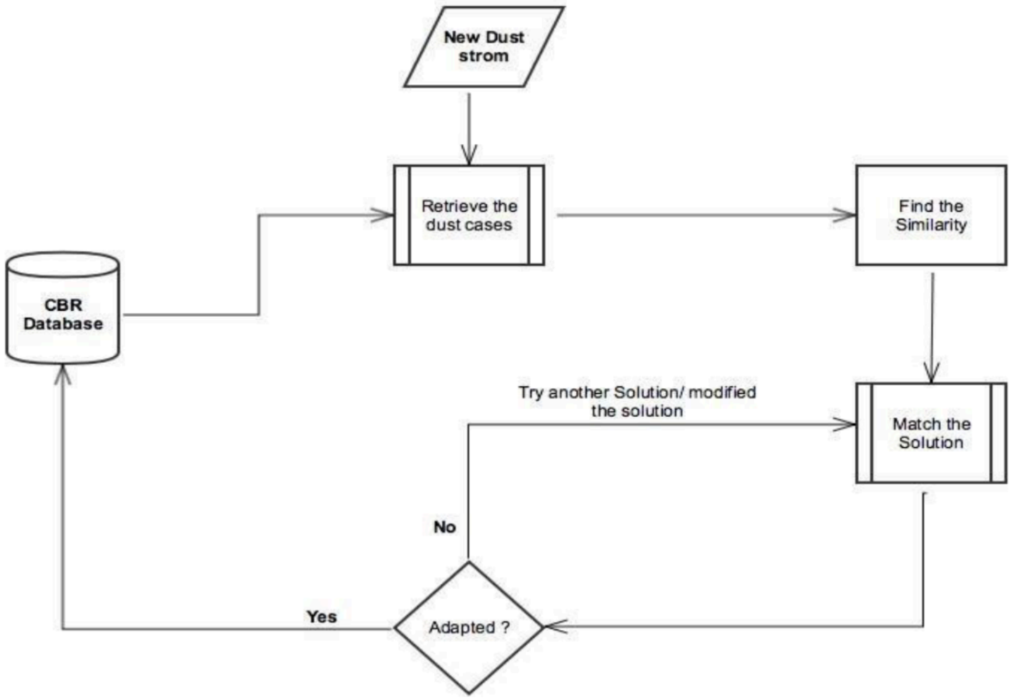


Figure 7. CBR cycle in stage two (Al Murayziq, Kapetanakis & Petridis, 2015).

Stage Three: Using RBS to Generate Safety Actions to Reduce the Impact of the Dust

The main idea that underpins the rule based system is that it offers the means to take the task-based knowledge of individuals and apply it in a computer. Once stored on the computer, that vast body of knowledge can be accessed by individuals whenever it is needed. It is quite possible that the computer will make inferences and draw conclusions based on that body of knowledge so that advice can be offered. Moreover, when possible, the computer will also demonstrate the logic behind the advice that it has offered. The empirical literature has established that rule based systems use the rules of IF-THEN, a set of facts and several controllers governing the application of the rules to provide the facts. The purpose of an expert system is to utilise knowledge but also encode that knowledge within a set of rules.

When the expert system and an expert are both provided with the same information, they will both work in the same way. One of the benefits of the rule based system is that it is relatively simple to apply and, therefore, is suitable for use in a range of applications. The two main ways in which RBS is applied are forward-chaining and backward-chaining. Problems such as that this work are solved using forward-chaining whereby inputs are used to solve the problem by making a prediction (i.e. the dust key features are used to arrive at the solution). It is the prediction that is the output in this scenario. The system is data driven and compares data in its working memory with the state of weather parameters (IF parts) of the rules and identifies the rules to fire. Forward-chaining should be selected when all of the actions are already known and the purpose of the system is to establish where the data goes to. For the purpose of the current study, the action is known because primary actions have been collected from the meteorology department and experts in the field.

Before RBS can be used, it is necessary to identify and evaluate the rules. This is achieved using the Chi-squared Automatic Interaction Detector (CHAID) algorithm which is a form of decision tree and has been extensively used in the empirical literature to support decisions and classify problems. CHAID is preferred to the alternative algorithms because it has been found to generate more rules than the alternatives such as the C&R tress, Apriori and C5.0. Moreover, the vast majority of these rules have been extensively evaluated by weather experts.

Chi-squared formula:

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{4}$$

CHAID produced a set of rules based on the weather status in the old dust base and these rules were sent to weather experts so that they could evaluate them using their own expertise and experience. In order for RBS to become operational, the C# language must be used to code the rules and actions. Using the forward chain, it will then be possible to forecast future dust storms from the RBS side and advise on the best course of action to minimise the harm to health and property. The accuracy of the prediction could be improved by combining the prediction with the BN-CBR forecast.

In the unlikely event that the RBS prediction and BN-CBR forecast conflicts, priority will be given to the BN-CBR forecast because this could use the 5 or 7NN before arriving at a final result. Moreover, all of the actions produced and the predictions in RBS can be stored in the dust base in order to expand the history of the dust base and enable it to be used in future BN-CBR processes or further research studies.

In order to utilise RBS in a dust project, the initial step is to take the forecasted dust storm into consideration and generate dust rules using the algorithm. The rules generated will be based on all dust cases and then safety actions will be ranked from high to low risk. Any action taken will comply with the relevant safety standard. Once these are prepared, the anticipated weather values are entered and RBS matches the rules with the actions. These actions will help to solve the problem but also to combine the dust cases with the solution in order to accumulate a compendium of dust events for future use.

The C# language is used to program the steps. There are two ways in which RBS can be constructed. The first approach involves the widely-used CLIPS tool which has been shown in the empirical literature to be highly efficient and easily maintained. However, one notable downside associated with CLIPS is that all of the facts and rules have to be entered separately. When a query is made, the tool finds each fact that matches the rules for the purposes of execution. Moreover, the user must deal with the interface and, therefore, each query must be types in order to generate the relevant fact.

The second method is to programme the process of RBS using a suitable language. This work, the C# language has been selected using Visual studio (VB) because the user must engage with the interface to enter the rules and actions and generate the results. VB utilises the IF-THEN function that is used to enter rules in the IF statement and the actions in the THEN statement. That is considerably easier than the CLIPS command. Crucially, the researcher has prior knowledge of the C# language and this will help the process. A CRYSTAL function is used in VB to export the results and display the action. The CRYSTAL function helps to prepare and generate good worksheet documents with results and actions that can be applied on a particular day to help minimise the effects of dust events.

By using CRYSTAL reports, the user is able to graphically design data connections and the layout of reports. Furthermore, the Database Expert enables users to choose and link tables from a variety of data sources such as Microsoft Excel spreadsheets, Microsoft Access databases, Microsoft SQL server databases, Oracle databases, Business Objects Enterprise business views, and local file-system information. When designing reports, fields from these sources can be placed on the report design surface and they can also be used to create custom formulas (using BASIC or CRYSTAL's own syntax) to be placed on the design surface. At various stages when compiling a report, formulas can be evaluated as specified by the developer.

There is a large selection of formatting options that can be used with the different fields and formulas. Designers can utilise these formatting options either conditionally or absolutely. Furthermore, the data can be assigned to different bands and these can then be subdivided and conditionally suppressed if necessary. CRYSTAL reports support the use of sub-reports, graphs and also enable GIS functionality to a limited extent.

```
private void button1_Click_2(object sender, EventArgs e)
{
if (text_wind_speedop1.Text.Trim() == "")
{
text_wind_speedop1.Text = " ";
}
if (text_wind_speedop2.Text.Trim() == "")
{
text_wind_speedop2.Text = " ";
}
if (text_wind_speed1.Text.Trim() == "")
{
text_wind_speed1.Text = "0";
}
if (text_wind_speed2.Text.Trim() == "")
{
text_wind_speed2.Text = "0";
}
if (combmonthop1.Text.Trim() == "")
{
combmonthop1.Text = " ";
}
```

Figure 8. Part of IF-Then code.

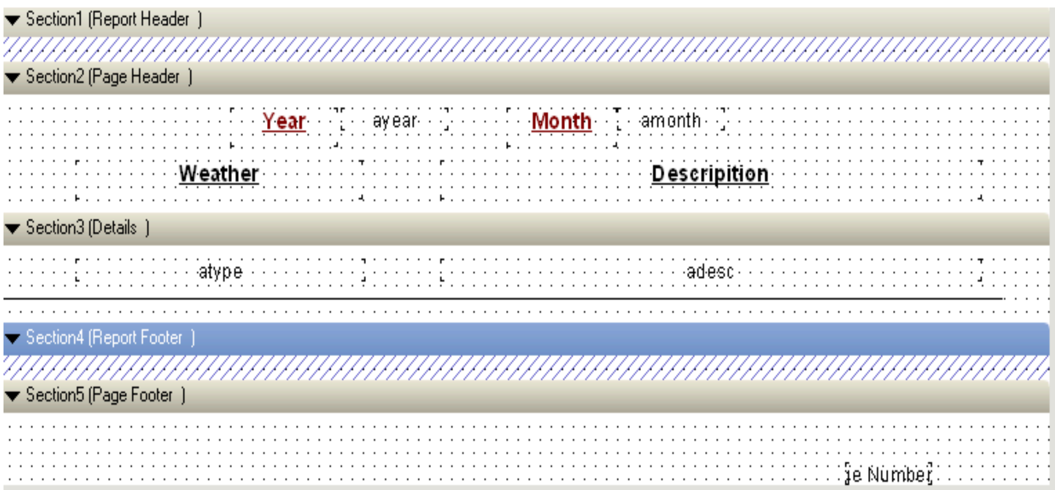


Figure 9. Crystal workbench.

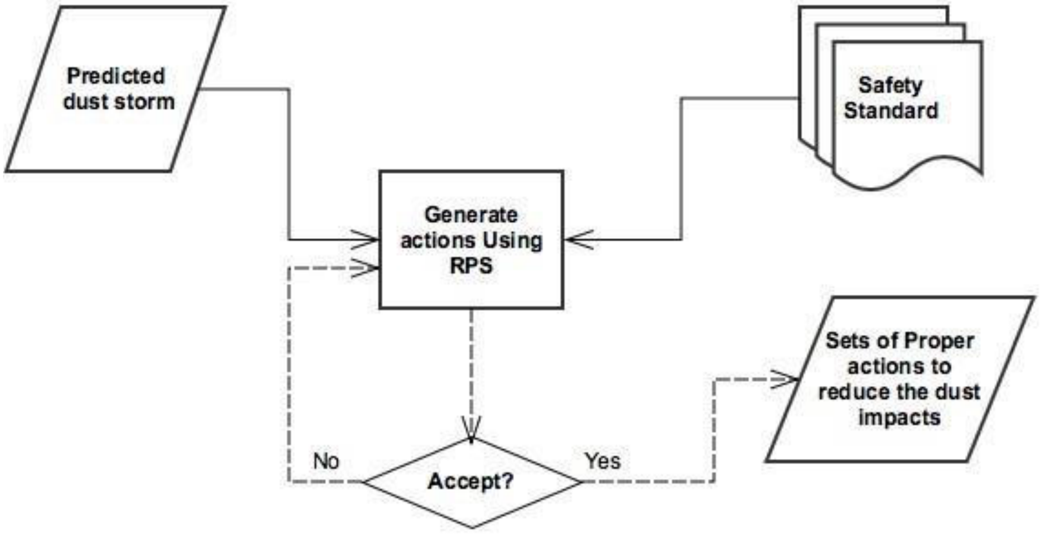


Figure 10. Applying RBS in stage three (Al Murayziq, Kapetanakis & Petridis, 2015).

Algorithm

Algorithm 1: (BN-CBR)-RBS Classify, Retrieve and Prediction

Input: Dust key attributes.

Output: A final prediction of new weather cases.

1. Apply BN algorithm to the old dust cases to re-categorise poorly classified cases.
2. Extract the BN dataset with new classification.
3. Compute the distance between the cases using Euclidean distance and calculate the similarity.
4. Match the new case with the closest old cases.
5. Use 3NN to come up with an accurate prediction.
6. Count the 3NN results as a final prediction of the new case.
7. While retaining the dataset and saving the new cases with predictions do the following:
8.   Apply CHAID algorithm to old dust cases for RBS stage.
9.   Extract the rules and adjust them.
10.   Store the safety action that suits to each dust category.
11. End.
12. Execute the rules with a new dust case.
13. List the state of the new weather case with predictions and proper actions.

From the above algorithm it is possible to predict forthcoming dust storms based on old dust cases. The input is as determined before from the key dust factors including wind, pressure …. etc. and the output will be the final prediction for the new weather case. Lines 1 to 6 describe the combining between BN-CBR in order to solve poorly classified cases and give the prediction based on the CBR cycle. Lines 7 to 13 show how RBS can work in parallel with BN-CBR to use the dust dataset to generate rules and give proper actions to help relieve the effects of forthcoming dust events.

Chapter Four

Experiments and results

Introduction

In this study the approach proposes how previous dust storm events can assist in forecasting new dust storms as well as the potential actions that should be taken to mitigate their potential impact. To illustrate this concept, a hybrid approach is proposed using a combination of BNs with CBR which indirectly estimates the similarity between previous storm cases and any new ones. Our approach reveals that BN-CBR may have the potential to be used for successful prediction and the possible mitigation of extreme weather events.

Dust Storm Characteristics

It is only by identifying the main features of a dust storm that it is possible to establish the inputs. Based on the available evidence, the following variables were identified as being the most significant features of dust storms. These features have been identified using the ranker algorithm and expert ranking, the wind direction does not take in consideration, because this study more focus on prediction dust event, that happened on cities that located near to the desert, more than monitor the path of the dust storm:

Month	Month of the recorded dust case.
Rainfall	How many mm of rain on that day.
W.S.M	Wind speed max.
P.M	Pressure max.
P.MIN	Pressure min.
P.S.M	Pressure sea level max.
P.S.MIN	Pressure sea level min.
H.M	Humidity max.
H.MIN	Humidity min.
A.T.M	Air temperature max.
A.T.MIN	Air temperature min.
S. A	Surface ability.

Importance of Dust Characteristics

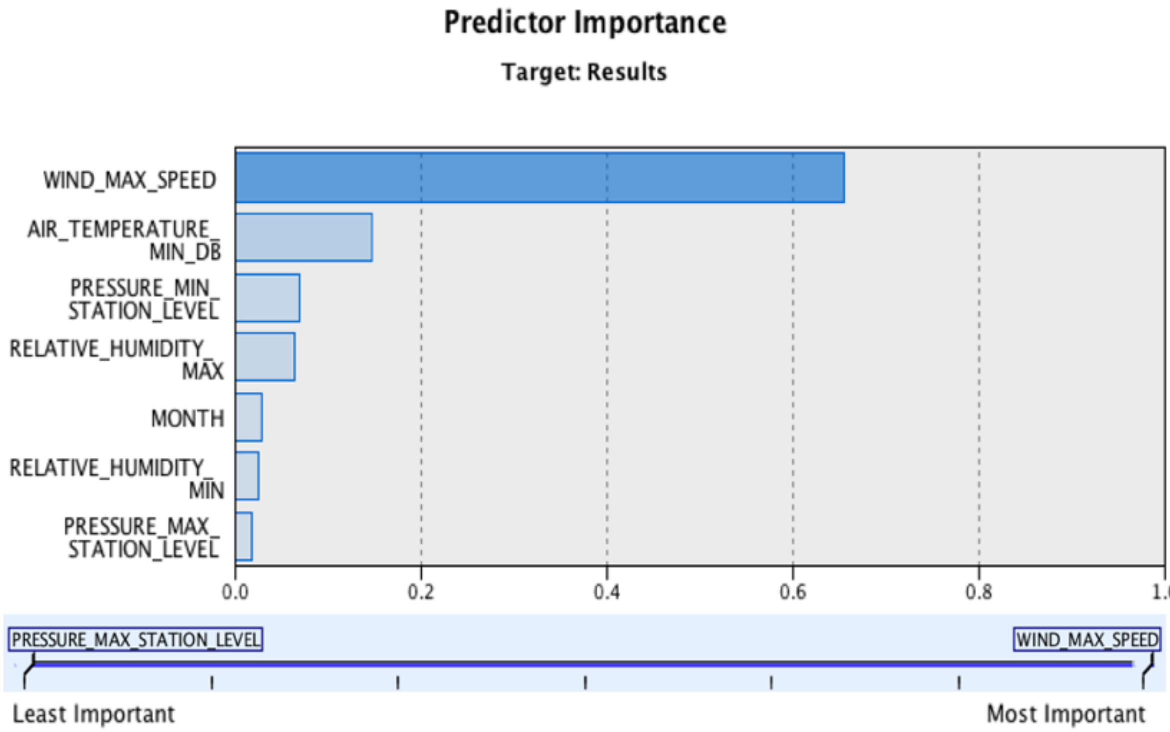
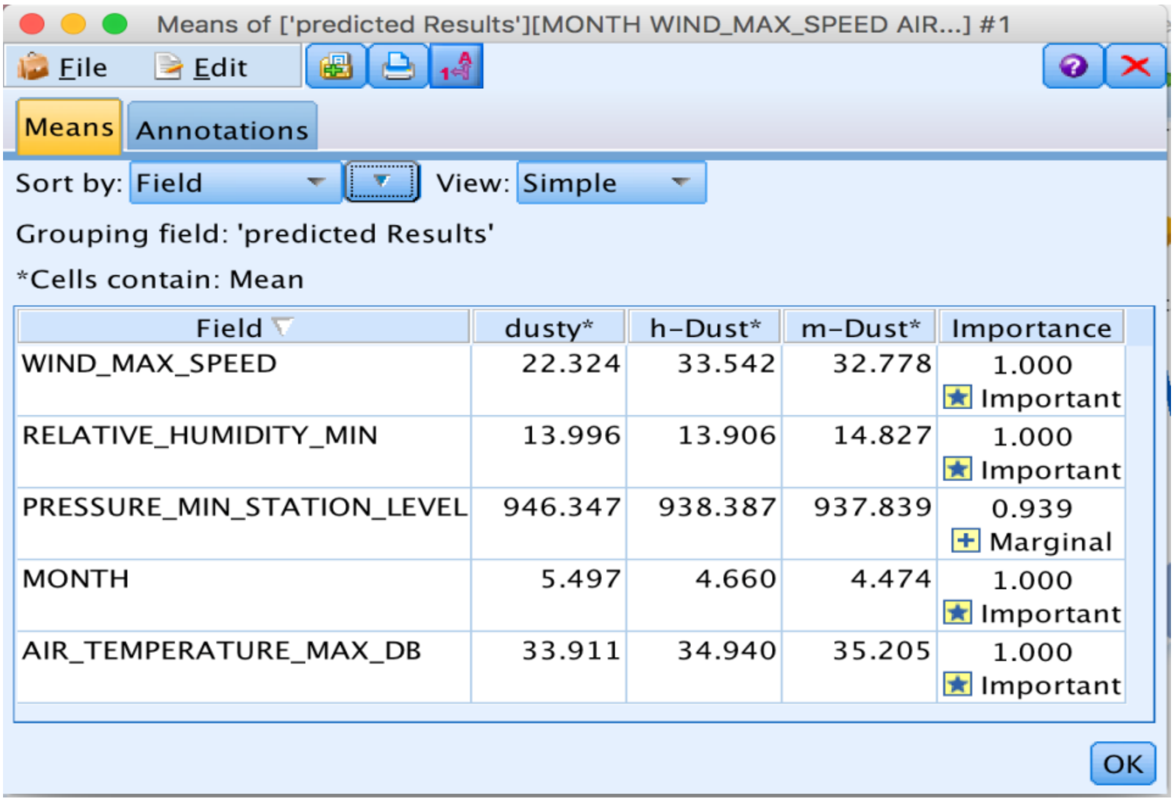


Figure 11. CHAID predictor importance (Al Murayziq, Kapetanakis & Petridis, 2017).

Figure 11 illustrates the CHAID algorithm for the Riyadh database. In this example, the target has been set into the case status and the algorithm is applied in order to establish the importance of the various attributes. By doing so, it is found that humidity is the least important attribute, whereby wind speed has the greatest effect for the inland cities.



Field	dusty*	h-Dust*	m-Dust*	Importance
WIND_MAX_SPEED	22.324	33.542	32.778	1.000 Important
RELATIVE_HUMIDITY_MIN	13.996	13.906	14.827	1.000 Important
PRESSURE_MIN_STATION_LEVEL	946.347	938.387	937.839	0.939 Marginal
MONTH	5.497	4.660	4.474	1.000 Important
AIR_TEMPERATURE_MAX_DB	33.911	34.940	35.205	1.000 Important

Figure 12. How dust features are important to each dust category with the mean.

As can be seen in the figure above, each dust attribute has a scale range to its optimal value in order to accrue the dust storm. Whenever these values are recorded from certain, it is more likely to record dust events in this weather status. This result has been established using CHAID algorithms for more than 7000 dust cases from the Riyadh station and using mean exporter.

Ranker Algorithm Results

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking

Attribute Evaluator (supervised, Class (nominal): 14 Dust ):
        Information Gain Ranking Filter

Ranked attributes:
 0.25602   2 WIND_MAX_SPEED
 0.04106   1 MONTH
 0.03267   4 PRESSURE_MIN_STATION_LEVEL
 0.02964   6 PRESSURE_MIN_SEA_LEVEL
 0.02408   8 AIR_TEMPERATURE_MIN_DB
 0.02406   9 RELATIVE_HUMIDITY_MAX
 0.01976   5 PRESSURE_MAX_SEA_LEVEL
 0.0189    3 PRESSURE_MAX_STATION_LEVEL
 0.01823  10 RELATIVE_HUMIDITY_MIN
 0.01258   7 AIR_TEMPERATURE_MAX_DB
 0.00997  11 RAINFALL_TOTAL
 0        12 Soil ability
 0        13 SYN_OBSRVN

Selected attributes: 2,1,4,6,8,9,5,3,10,7,11,12,13 : 13
```

Figure 13. Dust storm attribute ranking.

Figure 13 presents the results of the ranking of dust storm features produced using the Weka tool. The information gain algorithm (5) was employed in order to arrive at the information gain relative to the stated class. The results confirm that wind speed has the most significant effect in terms of creating a dust storm but that does not mean that a windy day will result in a dust storm being experienced. Rather, further analysis indicates that dust storms occur when a number of different contributory factors occur at the same moment. The box above lists each of these features according to the scale of their influence starting with wind speed followed by month, pressure, humidity and temperature. For each attribute, a weight is assigned and it is this weight that determines the contribution that each component makes to creating a dust event. Similarly, this same weight could be used in the CBR cycle for the Euclidean equation in order to establish the similarity between cases. The month is significant because much of the Middle East experiences winds called ‘Albawarh’ during the months of May, June, July and early August. These winds are especially active during the day but conditions typically become calmer during the night. Importantly, the ranking of attributes is in accordance with that of the meteorological experts.

$$\text{InfoGain(Class,Attribute)} = \text{H(Class)} - \text{H(Class} \mid \text{Attribute)}. \tag{5}$$

Meteorological Professional Results

It is vital that the different attributes are ranked and weighted appropriately because this will have a significant impact on the final predictions that are generated. Dr Abdaullah al misind (professor assistant at the University of Quassaim and a consultant at the Saudi meteorology department) was asked to rank the dust characteristics. He remarked that dust events are one of the most complex meteorological phenomena that only occur when a combination o attributes occur simultaneously. As previously explained, a dust storm will not occur simply because it is windy if the other components are within their normal range.

He suggested that the most significant signal of an impending dust event was the combination of high wind speed and low pressure. In addition, he stated that high wind speeds in areas experiencing drought significantly increases the likelihood of a dust storm. However, those same conditions would not yield a dust event in a country that has a different landscape such as lush, green forests. The surface features have, therefore, been combined with the dust characteristics to give the surface ability of the region. Dr Almisind was also asked to rank the dust features in terms of priority and his ranking was broadly similar to that of the current study. Using both rankings (the weather experts rank, the algorithm ranker and the CBR observation) it is now possible to arrive at the dust storm characteristics.

1	**Wind Speed**
2	Pressure min (Station or sea)
3	Month
4	Surface ability
5	Humidity min
6	Temperature max
7	Rain

Figure 14. The weather expert (Dr Almisind) dust ranking.

Dust Attribute Correlation

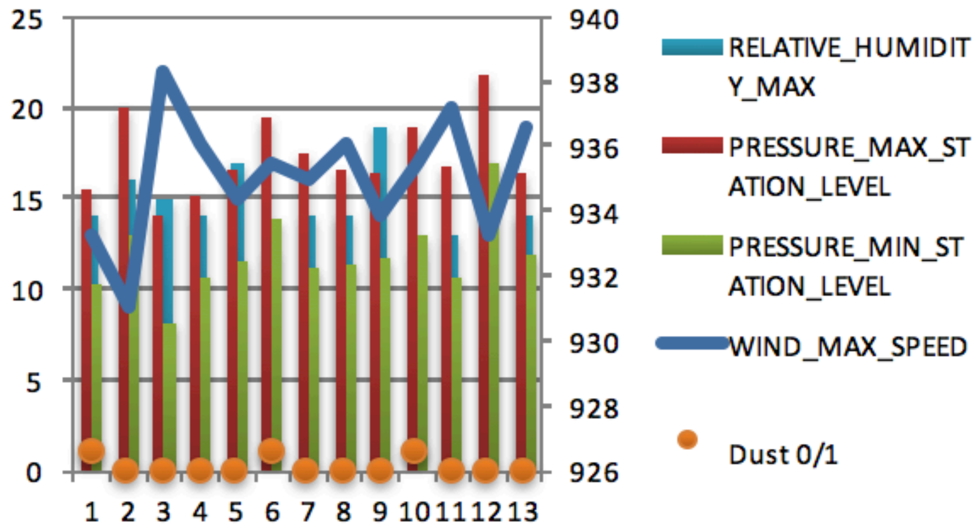


Figure 15. Dust storm attributes (Al Murayziq, Kapetanakis & Petridis, 2016).

As can be seen from Figure 15, a dust storm is likely to develop when a number of factors occur simultaneously: high wind speed, low pressure and low humidity. Another factor that increases the probability of experiencing a dust storm is high surface temperatures. As previously stated in the literature review, a dust event will only occur when more than one contributory factor happens in the same timeframe, does not matter which attribute occur first if not affected first by rain or high humidity. The diagram above provides an illustration of this and lends further support to the ranking that has been developed by the current study and also the weather expert. The above sample is tested for thirteen cases selected at random from the Riyadh dataset, whereby 0 indicates a non-dust event and 1 indicates a dust event. Three of the cases are dust cases and the remainder are non-dust cases. In some of the cases, all of the dust factors are present yet they are recorded as non-dust cases, possibly because the month or previous day has affected the weather status. For instance, had it rained on the previous day, the wind is unlikely to be able to carry the dust because wet surface is significantly more difficult to propel into the air.

Pressure vs Wind Speed

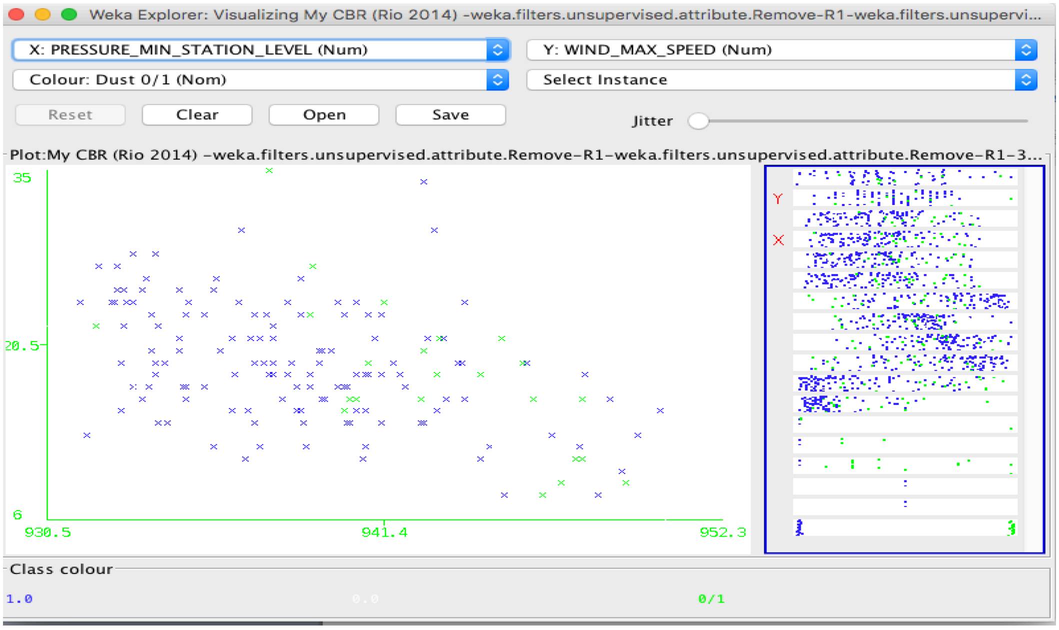


Figure 16. The correlation between pressure and wind from Riyadh dataset (Al Murayziq, Kapetanakis & Petridis, 2016).

Figure 16 illustrates the correlation between wind speed and air pressure. The combination of high wind speed and low air pressure is significantly more likely to result in a dust storm. More than 300 cases for Riyadh during 2014 are used as a test sample to demonstrate how these two factors are correlated. The minimum wind speed is given as 10km/h which indicates that wind speeds in excess of this could generate a dust storm. Similarly, a pressure of less than 941.4 mB could result in a dust storm. Referring to the figures provided, it is apparent that the majority of dust cases occur when these conditions are met but it is notable that there are a number of dust cases that lie outside of this range. This could possibly because the dust storms originated elsewhere but have travelled to Riyadh. Indeed, this was cited by the meteorology expert as one of the reasons why dust storms occur.

Pressure vs Humidity

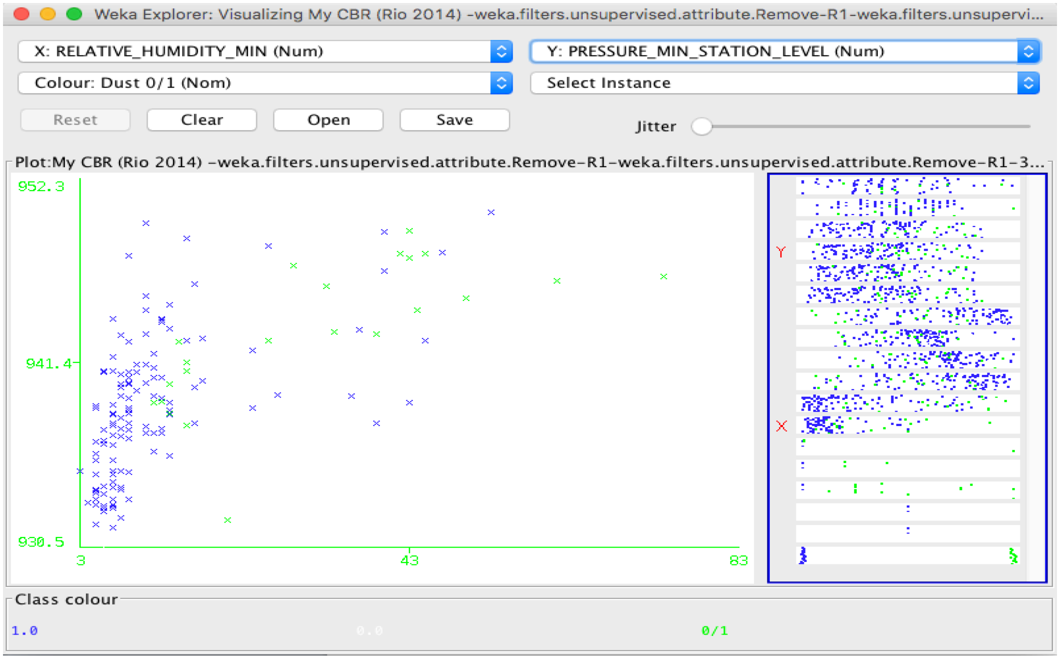


Figure 17. The correlation between pressure and humidity from Damam dataset.

The correlation between air pressure and humidity is presented in Figure 17 above. A combination of low air pressure and low humidity significantly increases the probability of experiencing a dust storm. The data used in Figure 16 are taken from the Damam dataset. Unlike Riyadh which is located inland, Damam lies on the coast of the Gulf Sea and the decision to include Damam in the study was to help demonstrate the importance of humidity in influencing the likelihood of experiencing dust storms. It is evident from Figure 16 that the number of dust cases increases markedly when the pressure falls to below 941.4 and the humidity is below 20. If the atmosphere were humid, this would make it more difficult for dust to be raised from the ground into the air. In addition, greater humidity will result in the soil becoming moist, thereby making it less likely that the wind will be able to carry dust into the air. There are a small number of cases in which dust events occur even when air pressure and humidity are both at elevated levels but this may be because these storms originated away from Damam or they could be associated with seasonal winds.

Months vs Dust Events

Weather experts have noted that the majority of dust events occur at certain times of year and this could be due to the effect of seasonal winds. As previously referred to, between March and May of each year, the north eastern region of Saudi Arabia (including Damam) experiences winds from the north east that are referred to as the Al Bawarh. Given the surrounding desert terrain, the majority of dust events that occur in these cities coincide with the Al Bawarh winds between March and May. This phenomenon was also referred to by Dr Almisind. Elsewhere in Saudi Arabia there are other seasonal winds that cause similar dust events. While it is certainly true that dust events do occur outside of these months when there is low pressure or winds that cause heavy cloud, it is worthwhile mentioning the significance of these seasonal winds.

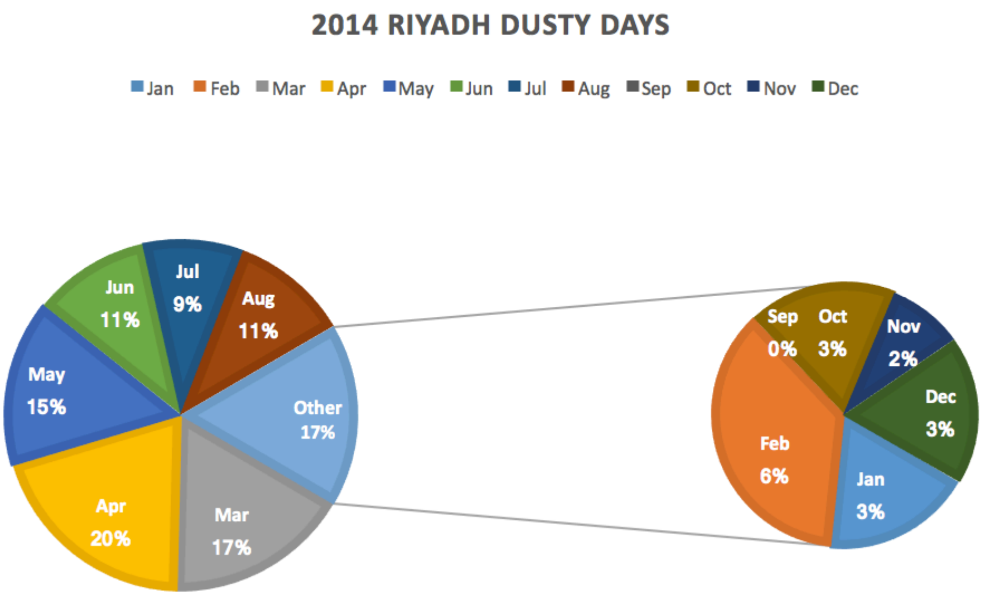


Figure 18. Dusty days in Riyadh city in 2014.

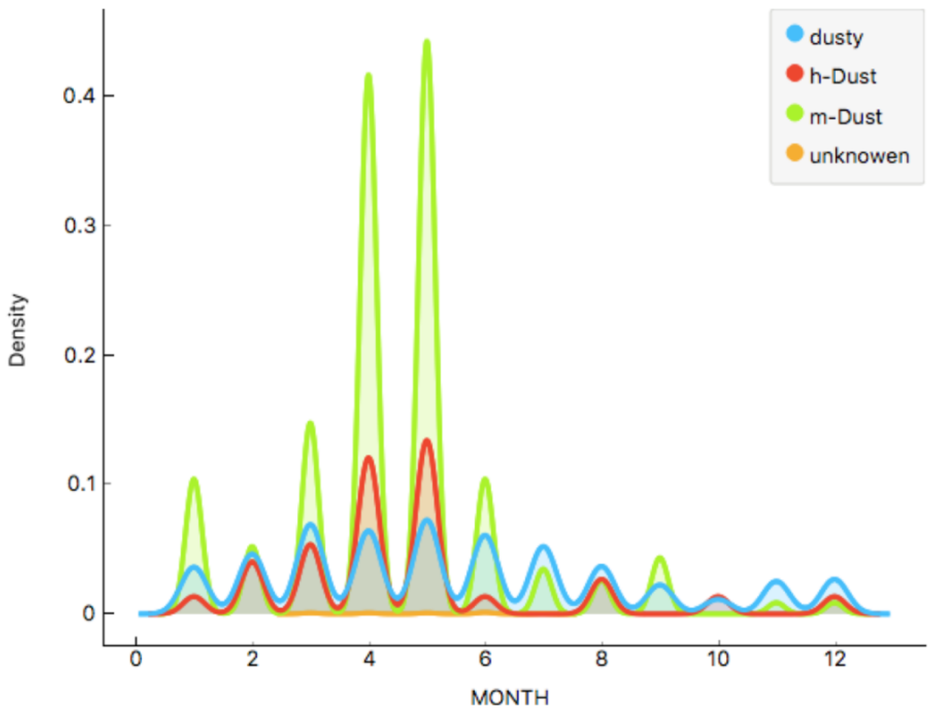


Figure 19. Dust cases distribution over month.

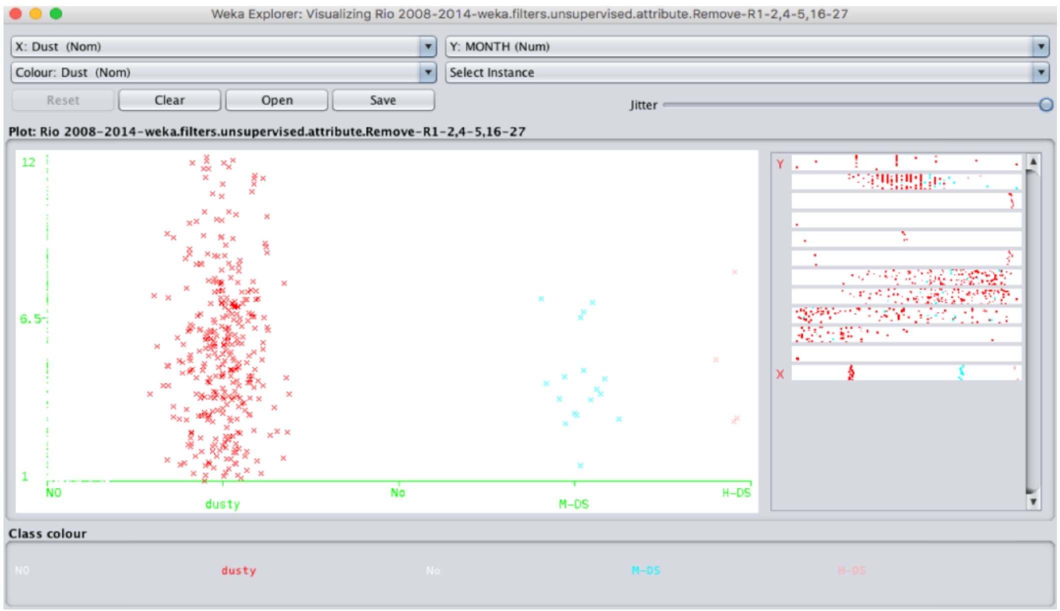


Figure 20. Riyadh 2013 dust cases distribution over the months.

Figures 19, 20 illustrates that most of dusty cases, medium dust and heavy dust happened between January and June.

Combining BN with CBR

The choice of Bayesian probability as a methodology is appropriate when it is necessary to reason and there are uncertain statements. More specifically, Bayesian methods are chosen when there is a need to develop an appreciation of a state of knowledge. When using this approach, a scale ranging from zero to one is employed to indicate the degree of confidence associated with a particular suggestion (Kapetanakis, 2012).

$$\mathrm{P}\ (\mathrm{E}|\mathrm{F}) = \frac{\mathrm{P}(\mathrm{F}|\mathrm{E})\ \mathrm{P}(\mathrm{E})}{\mathrm{P}(\mathrm{F})}$$

Where $P(E) \neq 0$ and $P(F) \neq 0$

A Bayesian network is employed in the current research in order to classify historical dust storms and categorise these events in terms of their severity. In the majority of cases, the storms are classified according to the code applied in the dataset. However, it is important to note that numerous cases have been mis-recorded or incorrectly classified and the Bayesian network is used to resolve this matter by re-classifying the data. Dust storms are assigned to one of three categories according to the strength of the storm as outlined below:

- **Dusty** – Dust is raised off the ground in the immediate vicinity of the weather station but without a dust whirl being generated and this does not result in a full-blown dust storm. While this does not constitute a dust storm, it could still pose a threat to human health and limit the ability of humans to operate outdoors. This is the most frequently reported condition in inland cities such as Riyahd. Furthermore, Dr Almisind notes that this is the most frequently experienced condition to afflict Middle Eastern countries.
- **Mid-dust** – A moderate dust storm that has increased in intensity over the course of the previous hour. Mid-dust occurs when there is a combination of low pressure, high wind speed and high temperature. These dust events could hinder agricultural processes and prevent humans from enjoying activities such as playing football. Furthermore, these events usually occur between seasons and in early Autumn.
- **Heavy dust storm** – A full dust storm that has increased in intensity within the previous hour. Heavy dust storms occur when there is very low pressure and high wind speed and are considered to be the most severe type of dust storm. Storms of this nature typically occur when there is extremely low pressure affecting a large land mass or several countries. In these cases it is quite possible for dust t be carried over many hundreds or

even thousands of miles. When heavy dust storms are forecast, preparations must be made in the interests of human safety owing to the large volume of dust that is being carried on the wind. Storms of this nature are common in Saudi Arabia, especially in inland cities such as Riyadh.

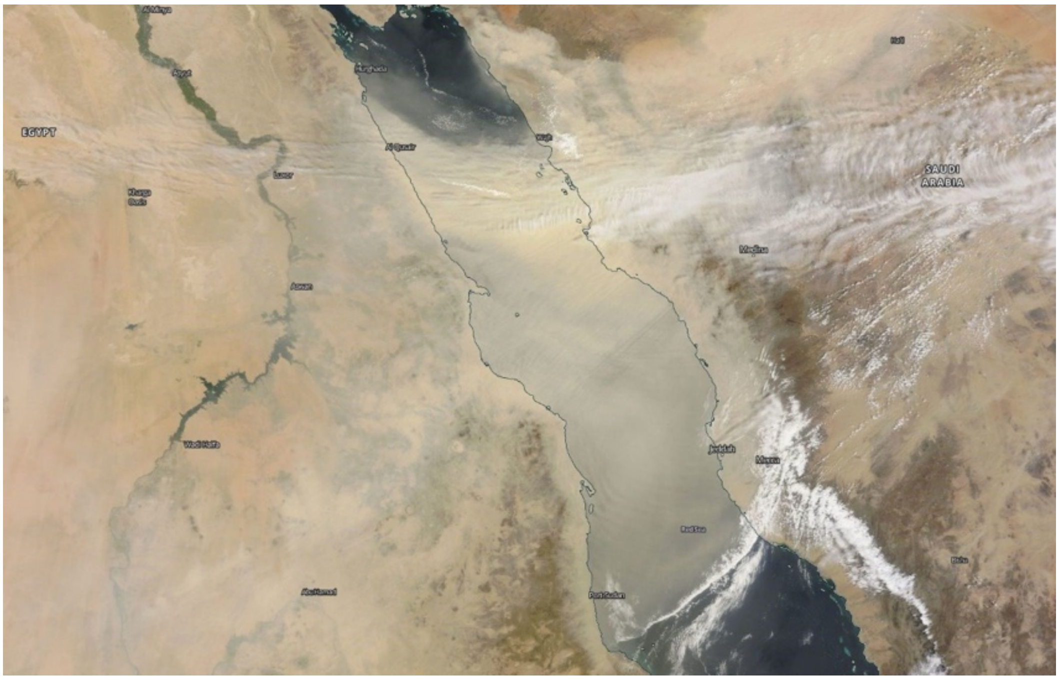


Figure 21. Dust storm transfers from one country to another. (2013, Jan). [digital image]. Retrieved from https://earthobservatory.nasa.gov/NaturalHazards/view.php?id=80112

Figure 22. Heavy dust storm hit Riyadh in Saudi Arabia (2013, May). [digital image]. Retrieved from https://earthobservatory.nasa.gov/NaturalHazards/view.php?id=80112

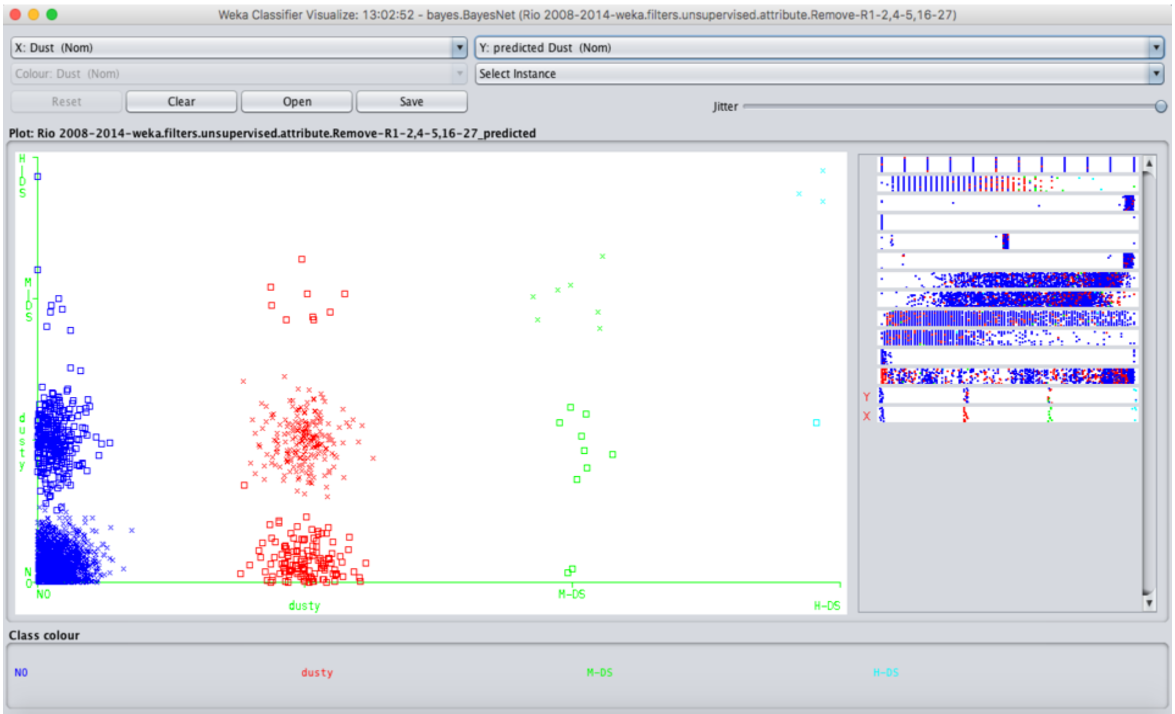


Figure 23. Using the BN to classify dust events

It can be seen from Figures 23 and 24 that the Bayesian network has performed well. The majority of cases have been classified and accurate predictions have been made for the categories of uncertain events. The decision to select the BN has been justified because it correctly classifies dusty, medium dust and heavy dust cases in the majority of cases with a low mean absolute error and satisfactory accuracy measurements in terms of precision and recall. The classification performance of alternative algorithms is below that of the BN in terms of accuracy.

Only a very small number of cases were seemingly mis-classified. These typically exhibit the main characteristics of dust even though that is not the case. These exceptions often follow a rainy day or occur outside of the dust season and these cases should be reclassified to the predicted status, to make sure they are in the correct status. Doing so helps to increase the accuracy of predictions. For the Riyadh dataset, a cross validation test is used to assess the performance of the algorithm.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       2081        81.3208 %
Incorrectly Classified Instances      478        18.6792 %
Kappa statistic                        0.3849
Mean absolute error                    0.1116
Root mean squared error                0.2708
Relative absolute error               88.7257 %
Root relative squared error          108.1556 %
Total Number of Instances             2559

=== Detailed Accuracy By Class ===

        TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC
Area  Class
        0.852    0.366    0.932      0.852   0.890      0.415  0.817     0.959   NO
        0.595    0.148    0.391      0.595   0.472      0.379  0.797     0.424   dusty
        0.353    0.006    0.300      0.353   0.324      0.321  0.933     0.259   M-DS
        0.750    0.001    0.600      0.750   0.667      0.670  0.888     0.751   H-DS
 W. Av  0.813    0.333    0.853      0.813   0.829      0.409  0.815     0.881

=== Confusion Matrix ===

   a    b    c    d   <-- classified as
 1863  317    7    0 |    a = NO
  134  209    7    1 |    b = dusty
    2    8    6    1 |    c = M-DS
    0    1    0    3 |    d = H-DS
```

Figure 24. BN classification result.

BN Node Structure

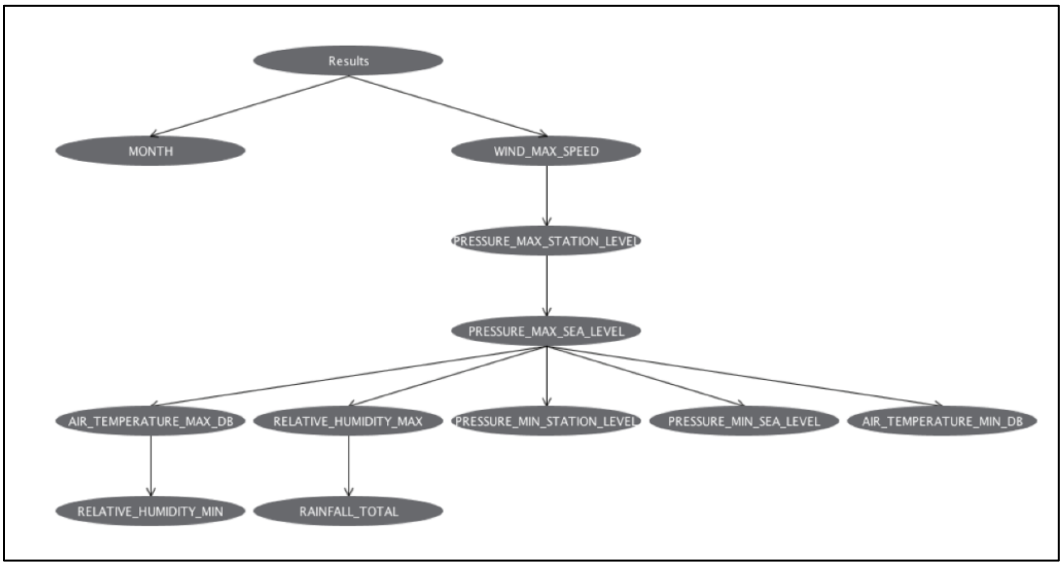


Figure 25. Standard BN structure.

Figure 25 above presents the tree structure produced when applying data for Riyadh to the BN standard using the K2 algorithm. The structure can be used to explain how the algorithm is able to work through the dust cases to provide logical explanations for each weather status. Of all the features, only the month vector is not correlated. The month vector has the importance of frequent dust events that occur when temperatures are high and the surface is experiencing drought conditions. Therefore, this structure is eliminated. The Markov Blanket structure tree provides a greater number of relationships between dust features, especially the in-wind speed node. However, this model also omits the month feature because the algorithm seeks out the highest score in K2 and in the Markov algorithm seeks the minimum set of variables conditionally according to the target.

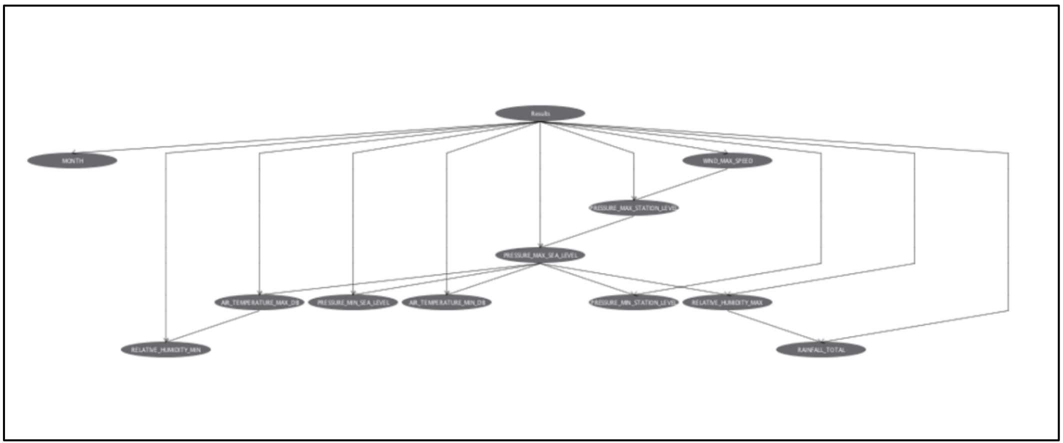


Figure 26. Markov node structure.

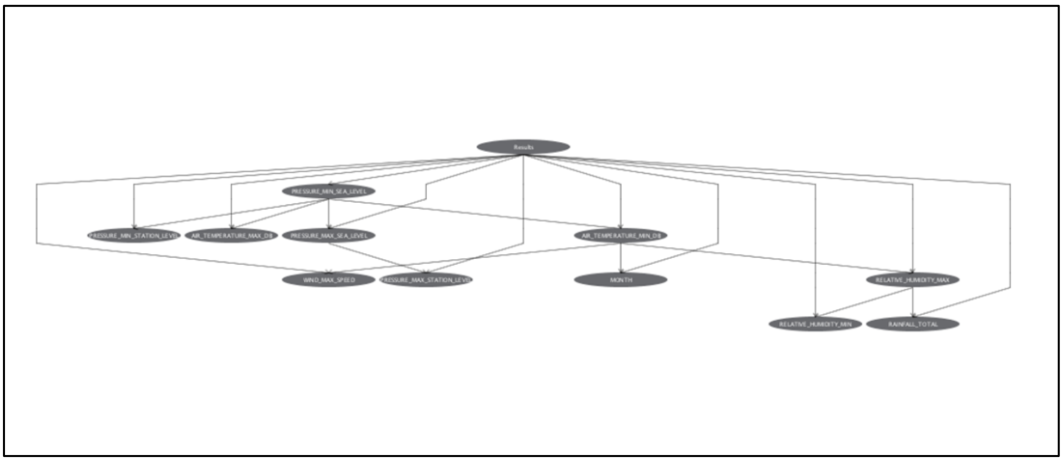


Figure 27. TAN BN structure.

In the TAN structure, the learning algorithm assumes that each attribute is dependent on its class and one other attribute from the feature set. In the ideal case scenario, the most realistic option is provided by the augmented naïve bays model owing to the fact that there is always a degree of correlation (albeit sometime limited) in the data. From these results it is apparent that the month attribute is related to another feature. The TAN structure provides a more reflective correlation between the components associated with dust events and effectively translates the weather expert's observations regarding such dust phenomena.

Bayesian Network Vs Lazy Learning

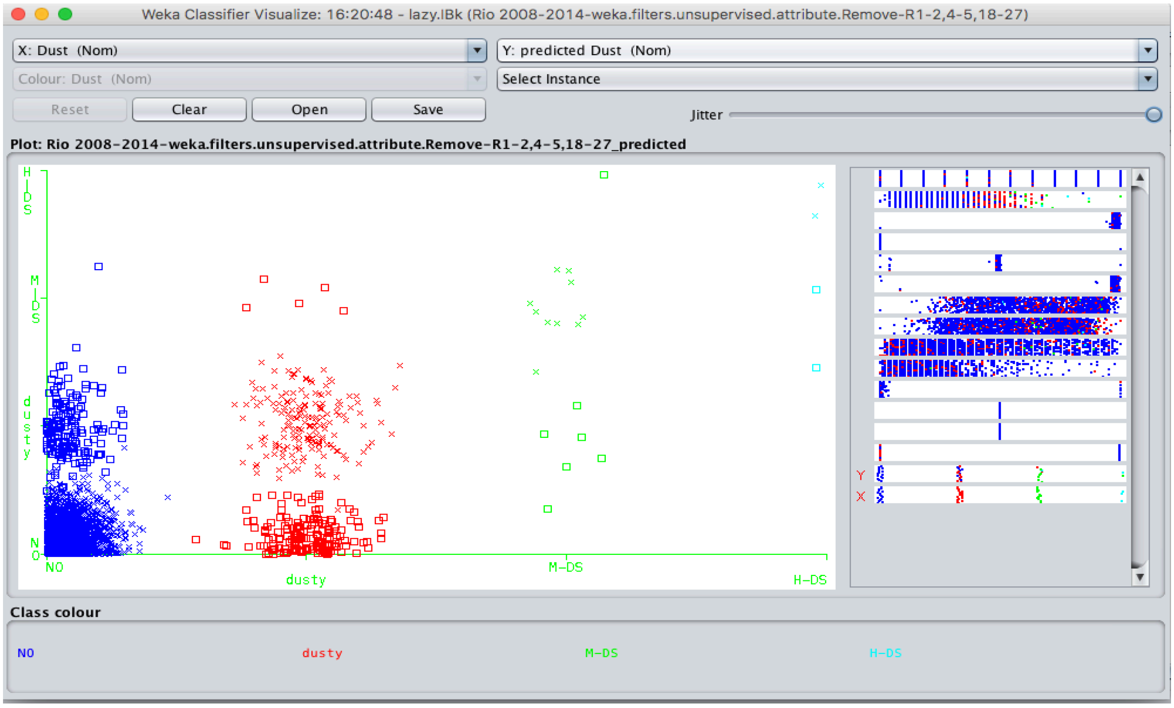


Figure 28. Applied lazy learning in Weka tool (Al Murayziq, Kapetanakis & Petridis, 2016).

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        2207              86.2446 %
Incorrectly Classified Instances       352              13.7554 %
Kappa statistic                          0.452
Mean absolute error                      0.0693
Root mean squared error                  0.262
Relative absolute error                 55.0904 %
Root relative squared error            104.6539 %
Total Number of Instances             2559

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC     ROC Area  PRC
               0.922    0.454    0.923      0.922   0.922      0.467   0.734     0.918   NO
               0.507    0.079    0.506      0.507   0.506      0.428   0.714     0.325   dusty
               0.588    0.003    0.556      0.588   0.571      0.569   0.784     0.330   M-DS
               0.500    0.000    0.667      0.500   0.571      0.577   0.775     0.418   H-DS
Weighted Avg.  0.862    0.399    0.863      0.862   0.863      0.463   0.731     0.832

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2017  169    1    0 |   a = NO
  168  178    5    0 |   b = dusty
    1    5   10    1 |   c = M-DS
    0    0    2    2 |   d = H-DS
```

Figure 29. Lazy learning classification result (Al Murayziq, Kapetanakis & Petridis, 2016).

Figures 28 and 29 demonstrate that dusty events can be reliably categorised using lazy learning. In addition, the figures present the confusion matrix. However, the figures also confirm that lazy learning is not able to reliably categorise mid-dust or heavy dust storms. Given that the current study is primarily concerned with mid-dust and heavy dust storms, it cannot be recommended that lazy learning should be applied in this research. In contrast, the BN was able to reliably categorise all types of dust events with a high degree of accuracy. Therefore, despite lazy learning achieving a high percentage in classification instances, it is not the preferred choice of algorithm.

BN Vs ANN

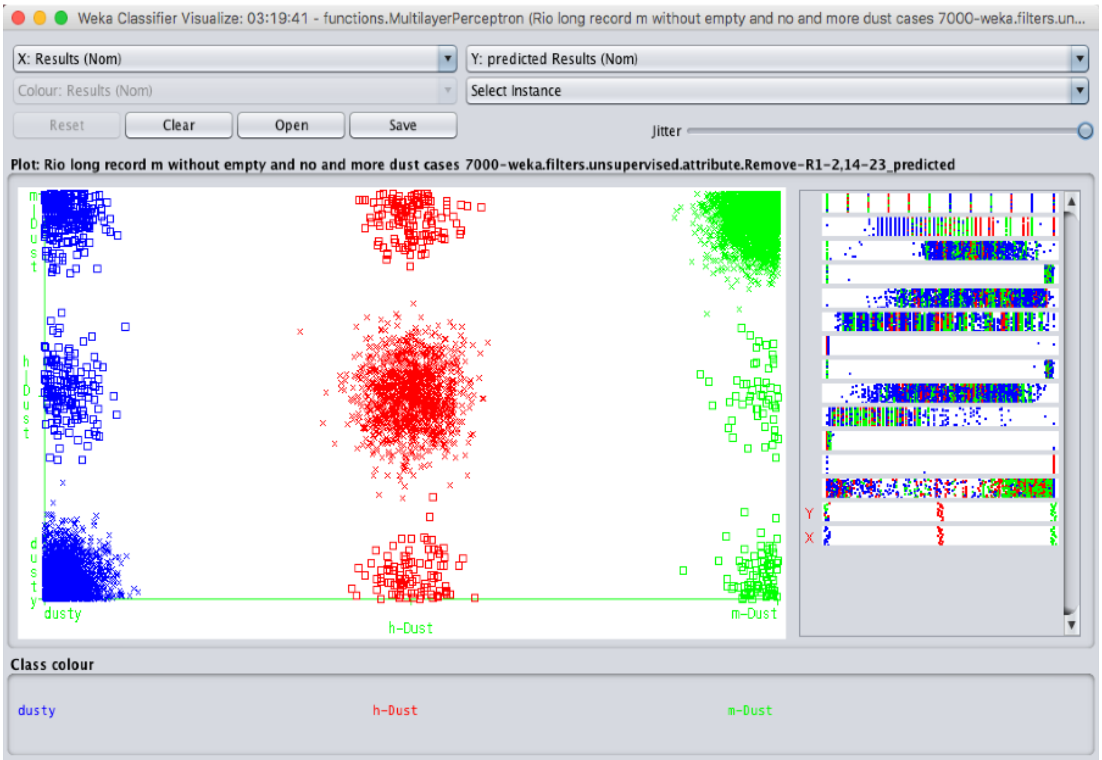


Figure 30. Applying ANN using Weka too.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        6868               89.5787 %
Incorrectly Classified Instances       799               10.4213 %
Kappa statistic                          0.8298
Mean absolute error                      0.0899
Root mean squared error                  0.2473
Relative absolute error                 21.7567 %
Root relative squared error             54.4135 %
Total Number of Instances             7667

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.806    0.033    0.913      0.806    0.856      0.803    0.917     0.891     dusty
                 0.856    0.034    0.862      0.856    0.859      0.824    0.923     0.834     h-Dust
                 0.965    0.108    0.900      0.965    0.932      0.860    0.950     0.921     m-Dust
Weighted Avg.    0.896    0.071    0.896      0.896    0.894      0.836    0.935     0.894

=== Confusion Matrix ===

    a    b    c   <-- classified as
 1846  158  287 |    a = dusty
   96 1315  125 |    b = h-Dust
   80   53 3707 |    c = m-Dust
```

Figure 31. ANN results on Riyadh Dataset.

Figure 29 and 30 indicate that the ANN is able to reliably classify heavy dust storms and the backpropagation algorithm almost classifies them correctly, the results for dusty and medium-dust events are somewhat disappointing when compared to the BN. While ANN performs respectably, the BN remains the preferred choice on performance grounds. The BN classifies cases according to probability and similarity, whereas lazy learning seeks to fully identify the cases and subsequently classify them. Meanwhile, ANN appears to be prone to overfitting whereby the algorithm follows every single value with a detrimental effect on the accuracy of the results. As such, overfitting could conceivably result in the predictions being well wide of the true data points. Furthermore, it is often the case that the data are noisy or limited. Therefore, it is preferable to make exceptions for these values and under-fit if ANN is to be used in place of the BN. Having tested a number of different algorithms, it is considered that the BN offers the best option for the dataset being used in the current research study.

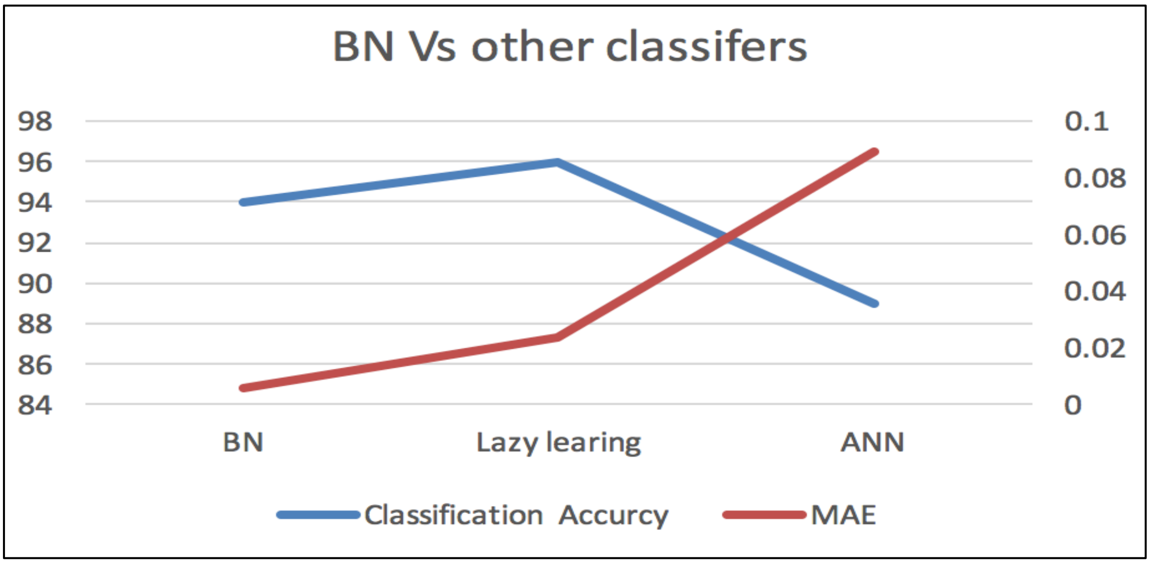


Figure 32. BNs vs other classification algorithms (Al Murayziq, Kapetanakis & Petridis, 2017).

Actual \ Predicted	dusty	h-Dust	m-Dust	unknowen	Σ
dusty	2140	35	116	0	2291
h-Dust	0	1536	0	0	1536
m-Dust	0	0	3840	0	3840
unknowen	0	7	3	0	10
Σ	2140	1578	3959	0	7677

Figure 33. Confusion matrix applying KNN with BN to the dust datasets.

Actual \ Predicted	dusty	h-Dust	m-Dust	unknowen	Σ
dusty	2138	29	122	2	2291
h-Dust	0	1536	0	0	1536
m-Dust	3	0	3837	0	3840
unknowen	5	0	0	5	10
Σ	2146	1565	3959	7	7677

Figure 34. Confusion matrix applying ANN to the dust datasets.

The figures above confirm that the BN outperforms the ANN and lazy learning in terms of accuracy and absolute error. These findings are based on more than 7,000 observations of dust events across a spectrum of dust categories. Case-based reasoning is used to match the old and new cases, thereby making it possible to forecast dust storms. In the methodology that has been selected for the current study, a benchmark is created based on historical examples and this makes it possible to gain a better understanding of dust events. The BN is employed to classify the cases and the following paragraph sets out the processes that are involved in this process.

The Weka tool is used to apply the BN algorithm to historical dust cases, thereby enabling the algorithm to identify, categorise and classify the dust cases to restrict the number of unclassified cases before the CBR cycle begins. This process could enhance the CBR forecast. It is also at this stage that the BN may produce its first prediction. When the

BN is applied with the dataset it classifies the dust cases and unclassified cases. In addition, it classifies the new weather cases. Based on this initial classification, it is possible that initial predictions could be arrived at that can subsequently be confirmed using CBR in the next stage of the process. Classification using BN is based on probability whereby the features of new weather developments are matched against historical cases. The BN algorithm has previously been found to reliably predict developments in other settings such as predicting road congestion in urban areas. One particular study of Beijing showed how a BN could be used to represent the stochastic nature of traffic congestion. The study also demonstrated how congestion in Beijing could effectively be reduced by simultaneously building more roads and developing the bus network. In addition, BN have been used to predict future movements in share prices. The first process is to determine the network based on daily stock prices and then applying the network to forecast future stock prices. These predictions have been found to be reasonably accurate. When using BN, the initial predictions are taken into account and must then be confirmed or rejected (Liu & Feng & Wang et al, 2014).

Tree augmented naïve Bayes (TAN) Vs CHAID Classification

CHAID was able to reclassify the dust cases relatively accurately but the BN was selected as the preferred option owing to the fact that it is able to compute the joint probability table between the variables, determine whether two variables are conditionally independent, and determine the distribution of non-evidence variables based on the evidence. Indeed, Janssens et al. note that BN is a better option than a decision tree. This work, these models performed broadly equally well.

Moreover, Janssens et al. 2014 also suggested that BN are better able to capture the complexity of the underlying decision-making process, bearing in mind the various (inter)dependencies between the different variables. In the current research study, the BN provides valuable insight into the way in which dust attributes, the time of year and the results are related. This relationship is most evident between the months of March and May when wind speeds pick up and pressure falls. This indicates that even if the BN and CHAID models use different variables, these variables are closely related. However, it is important to be aware that BN may associate variables in ways that are direct, indirect or complex, thereby making it more difficult to interpret. For example, this work, air temperature and pressure are associated with the target node and also with each other. In addition, they are related to two further variables that have an indirect bearing on wind speed. Meanwhile, decision trees are able to yield decision rules in a relatively simple process that provides a more direct interpretation tool (Janssens et al., 2004).

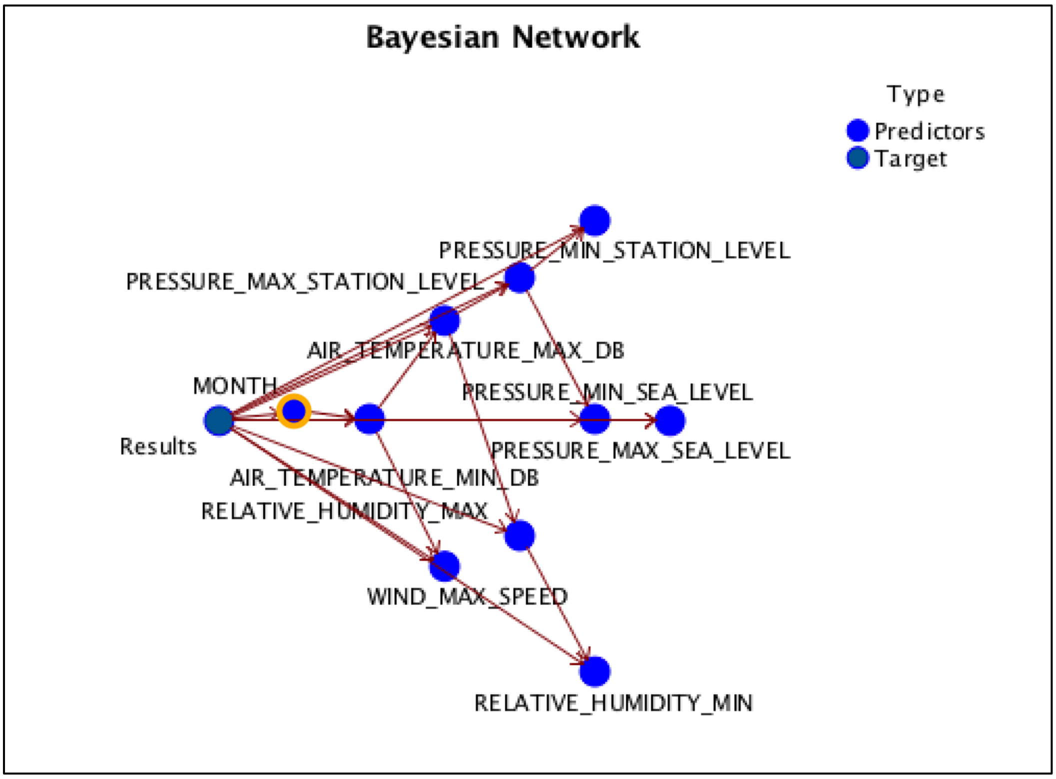


Figure 35. BN TAN structure.

Case Based Reasoning (CBR)

Once the BN process is complete, the CBR cycle will be initiated in order to forecast the weather status of new weather cases. This process relies on the BN classification being exported as an Excel spreadsheet. New cases are included in the BN results and can be chosen according to the date of the cases and subsequently imparted to the MyCBR tool so as to initiate the CBR process. Euclidean distance is used to measure similarity in the CBR process. The Euclidean distance equation provides an indication of the degree of similarity between the current case under consideration and the historical cases stored in the database. A separate Euclidean function is applied for each of the variables, but care must be taken to initialise the different variable weights.

Then the equation results for the different cases are aggregated in order to indicate the degree of similarity. The resulting values are then ranked in descending order. The degree of similarity between the current case and historical cases is established using the above formula. In addition, it provides a selection of cases that are similar to the current case being considered.

Figure 36 illustrates data being imported into the myCBR tool. This process involves using Euclidean distance equation measurements to establish the degree of similarity between cases. In order to reflect the particular features of the dust storm, the weights applied to some of the variables will need to be adjusted.

The weather status of historical cases was predicted using case-based reasoning which involves randomly selecting subsets from the database. Approximately 10-12% of the database is sampled and the classification of these samples remains unknown. Rather, case-based reasoning was used to classify them using similarity measures and 3NN classification. For each of these cases, ten separate experiments were performed and the results were averaged in order to estimate how efficiently CBR is able to forecast the weather status. The results indicate that the accuracy of CBR is 60-80% in pure CBR when the data are imported into myCBR without first performing the BN process (see Figure 36). However, Table 3 confirms that combining CBR with a BN classifier significantly improves the accuracy, yielding results of 80-90% when forecasting no dust, dusty or mid-dust storms. It is highly likely that this same technique could reliably forecast heavy dust storms but this cannot be confirmed without a much larger number of samples. The test has been conducted in both pure CBR and BN-CBR for seven, eight and ten folds to test whether forecast efficiency is maximised when using just CBR or CBR with BN.

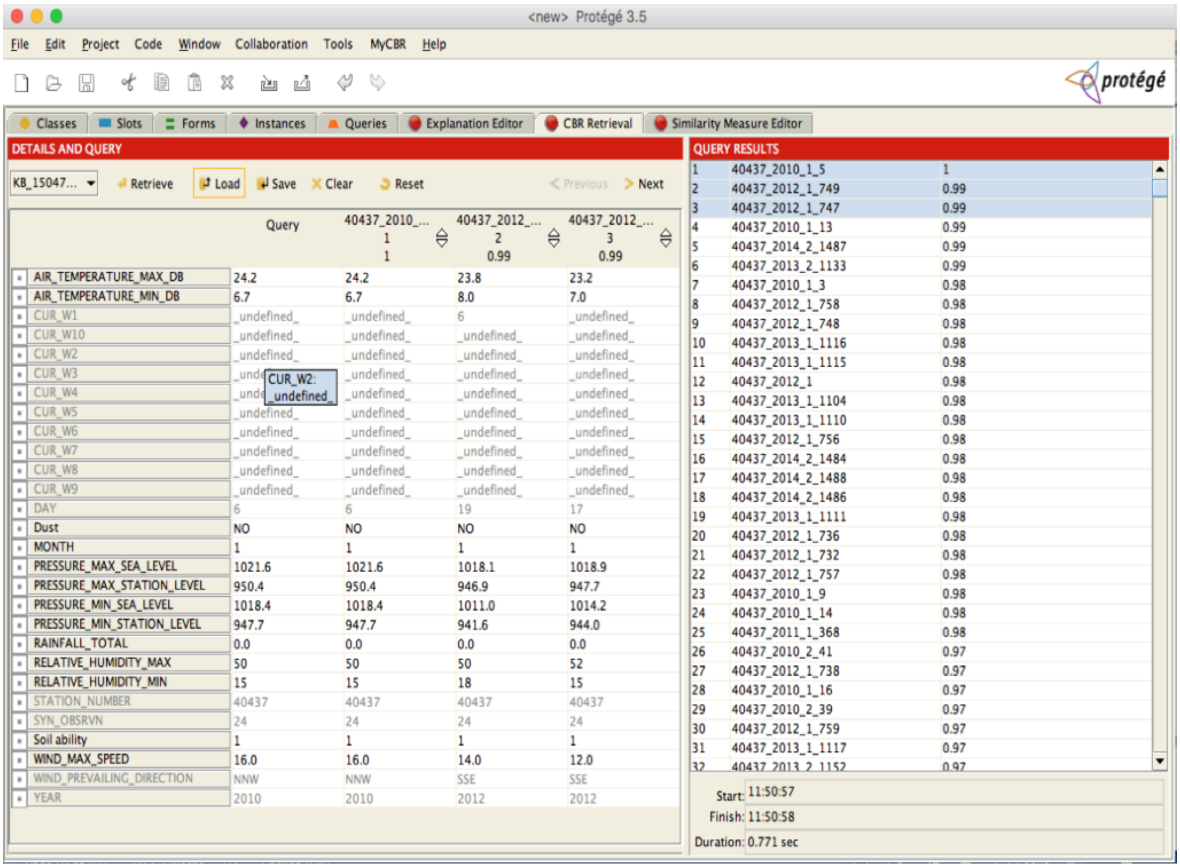


Figure 36. Similarity between one case versus other cases in MyCBR tool.

Table 3. BN-CBR result in 10 folds.

ID	Result of the case	Result of 1st 3NN	Result of 2nd 3NN	Result of 3rd 3NN	Result of 4th 3NN	Result of 5th 3NN	Result of 6th 3NN	Result of 7th 3NN	Result of 8th 3NN	Result of 9th 3NN	Result of 10th 3NN	Final Result
5-2546	No	No	No	No	No	No	No	No	No	No	No	No
3-1910	No	No	No	No	No	No	No	No	No	No	No	No
3-1787	M-D	M-D	M-D	M-D	dusty	M-D	M-D	dusty	dusty	M-D	M-D	M-D
7-2555	No	No	No	No	No	No	No	dusty	No	No	dusty	No
4-1023	M-D	M-D	M-D	M-D	M-D	M-D	M-D	M-D	No	M-D	dusty	M-D
5-255	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	M-D	dusty
8-2557	M-D	M-D	M-D	H-D	dusty	dusty	M-D	M-D	dusty	dusty	dusty	dusty
4-1299	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	No	dusty	dusty
2-510	M-D	M-D	M-D	M-D	M-D	M-D	dusty	dusty	No	M-D	dusty	M-D
2-453	No	No	No	No	No	No	No	No	No	No	dusty	No

The table above presents the results when applying ten folds for randomly selected dust cases taken from various categories. The nearest neighboured 3NN was applied and the results are presented in name cells. Overall, the results obtained using CBR with BN support are found to be 90% accurate. The folds has change to measure if there are improvement in accuracy or not.

Table 4. BN-CBR result in 9 folds.

ID	Result of the case	Result of 1st 3NN	Result of 2nd 3NN	Result of 3rd 3NN	Result of 4th 3NN	Result of 5th 3NN	Result of 6th 3NN	Result of 7th 3NN	Result of 8th 3NN	Result of 9th 3NN	Final Result
5-2546	No	No	No	No	No	No	No	No	No	No	No
3-1910	No	No	No	No	No	No	No	No	No	No	No
3-1787	M-D	M-D	M-D	M-D	dusty	M-D	M-D	dusty	dusty	M-D	M-D
7-2555	No	No	No	No	No	No	No	dusty	No	No	No
4-1023	M-D	M-D	M-D	M-D	M-D	M-D	M-D	M-D	No	M-D	M-D
5-255	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty
8-2557	M-D	M-D	M-D	H-D	dusty	dusty	M-D	M-D	dusty	dusty	Not good results
4-1299	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	No	dusty
2-510	M-D	M-D	M-D	M-D	M-D	M-D	dusty	dusty	No	M-D	M-D
2-453	No	No	No	No	No	No	No	No	No	No	No

The table above presents the results when applying nine folds for randomly selected dust cases taken from various categories. The nearest neighboured 3NN was applied and the results are presented in name cells. Overall, the results obtained using CBR with BN support are found to be 90% accurate. Case 8-2557 had a low level of accuracy and the 3NN did not perform well for the 9 folds.

Table 5. BN-CBR result in 8 folds.

ID	Result of the case	Result of 1st 3NN	Result of 2nd 3NN	Result of 3rd 3NN	Result of 4th 3NN	Result of 5th 3NN	Result of 6th 3NN	Result of 7th 3NN	Result of 8th 3NN	Final Result
5-2546	No	No	No	No	No	No	No	No	No	No
3-1910	No	No	No	No	No	No	No	No	No	No
3-1787	M-D	M-D	M-D	M-D	dusty	M-D	M-D	dusty	dusty	M-D
7-2555	No	No	No	No	No	No	No	dusty	No	No
4-1023	M-D	M-D	M-D	M-D	M-D	M-D	M-D	M-D	No	M-D
5-255	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty
8-2557	M-D	M-D	M-D	H-D	dusty	dusty	M-D	M-D	dusty	Not good results
4-1299	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty
2-510	M-D	M-D	M-D	M-D	M-D	M-D	dusty	dusty	No	M-D
2-453	No	No	No	No	No	No	No	No	No	No

The table above presents the results when applying eight folds for randomly selected dust cases taken from various categories. The nearest neighboured 3NN was applied and the results are presented in name cells. Overall, the results obtained using CBR with BN support are found to be 90% accurate. Again, case 8-2557 had a low rate of accuracy and the 3NN did not perform well for the 8 folds.

Table 6. BN-CBR result in 7 folds.

ID	Result of the case	Result of 1st 3NN	Result of 2nd 3NN	Result of 3rd 3NN	Result of 4th 3NN	Result of 5th 3NN	Result of 6th 3NN	Result of 7th 3NN	Final Result
5-2546	No	No	No	No	No	No	No	No	No
3-1910	No	No	No	No	No	No	No	No	No
3-1787	M-D	M-D	M-D	M-D	dusty	M-D	M-D	dusty	M-D
7-2555	No	No	No	No	No	No	No	dusty	No
4-1023	M-D	M-D	M-D	M-D	M-D	M-D	M-D	M-D	M-D
5-255	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty
8-2557	M-D	M-D	M-D	H-D	dusty	dusty	M-D	M-D	M-D
4-1299	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty
2-510	M-D	M-D	M-D	M-D	M-D	M-D	dusty	dusty	M-D
2-453	No	No	No	No	No	No	No	No	No

The table above presents the results when applying eight folds for randomly selected dust cases taken from various categories. The nearest neighboured 3NN was applied and the results are presented in name cells. In this scenario, the accuracy rate rose to 100% and case 8-2557 increased in accuracy 7-fold.

Table 7. Pure CBR result in 10 folds.

ID	Result of the case	Result of 1st 3NN	Result of 2nd 3NN	Result of 3rd 3NN	Result of 4th 3NN	Result of 5th 3NN	Result of 6th 3NN	Result of 7th 3NN	Result of 8th 3NN	Result of 9th 3NN	Result of 10th 3NN	Final Result
5-2546	No	No	No	No	No	dusty	No	dusty	dusty	No	No	No
3-1910	No	No	No	dusty	dusty	dusty	dusty	dusty	No	No	dusty	dusty
3-1787	M-D	M-D	M-D	dusty	dusty	M-D	dusty	dusty	dusty	No	dusty	dusty
7-2555	No	No	No	No	No	No	No	dusty	No	No	dusty	No
4-1023	M-D	M-D	M-D	M-D	M-D	dusty	dusty	M-D	No	M-D	dusty	M-D
5-255	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	No	dusty	No	dusty
8-2557	M-D	H-D	H-D	dusty	dusty	dusty	M-D	M-D	dusty	dusty	dusty	dusty
4-1299	dusty	dusty	dusty	dusty	dusty	No	No	No	dusty	dusty	dusty	dusty
2-510	M-D	M-D	M-D	M-D	M-D	M-D	dusty	dusty	dusty	M-D	dusty	M-D
2-453	No	No	No	No	No	No	No	dusty	dusty	No	No	No

The table above presents the results when applying ten folds for randomly selected dust cases to test the results of pure CBR against those of BN-CBR. The nearest neighboured 3NN was applied and the results are presented in name cells. These results indicate that CBR is able to predict the final results with an accuracy of 70%.

Table 8. Pure CBR result in 9 folds.

ID	Result of the case	Result of 1st 3NN	Result of 2nd 3NN	Result of 3rd 3NN	Result of 4th 3NN	Result of 5th 3NN	Result of 6th 3NN	Result of 7th 3NN	Result of 8th 3NN	Result of 9th 3NN	Final Result
5-2546	No	No	No	No	No	dusty	No	dusty	dusty	No	No
3-1910	No	No	No	dusty	dusty	dusty	dusty	dusty	No	No	dusty
3-1787	M-D	M-D	M-D	dusty	dusty	M-D	dusty	dusty	dusty	No	dusty
7-2555	No	No	No	No	No	No	No	dusty	No	No	No
4-1023	M-D	M-D	M-D	M-D	M-D	dusty	dusty	M-D	No	M-D	M-D
5-255	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	No	dusty	dusty
8-2557	M-D	H-D	H-D	dusty	dusty	dusty	M-D	M-D	dusty	dusty	dusty
4-1299	dusty	dusty	dusty	dusty	dusty	No	No	No	dusty	dusty	dusty
2-510	M-D	M-D	M-D	M-D	M-D	M-D	dusty	dusty	dusty	M-D	M-D
2-453	No	No	No	No	No	No	No	dusty	dusty	No	No

The table above presents the results when applying nine folds for randomly selected dust cases to test the results of pure CBR against those of BN-CBR. The nearest neighboured 3NN was applied and the results are presented in name cells. These results indicate that CBR is able to predict the final results with an accuracy of 70%.

Table 9. Pure CBR result in 8 folds.

ID	Result of the case	Result of 1st 3NN	Result of 2nd 3NN	Result of 3rd 3NN	Result of 4th 3NN	Result of 5th 3NN	Result of 6th 3NN	Result of 7th 3NN	Result of 8th 3NN	Final Result
5-2546	No	No	No	No	No	dusty	No	dusty	dusty	No
3-1910	No	No	No	dusty	dusty	dusty	dusty	dusty	No	dusty
3-1787	M-D	M-D	M-D	dusty	dusty	M-D	dusty	dusty	dusty	dusty
7-2555	No	No	No	No	No	No	No	dusty	No	No
4-1023	M-D	M-D	M-D	M-D	M-D	dusty	dusty	M-D	No	M-D
5-255	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	No	dusty
8-2557	M-D	H-D	H-D	dusty	dusty	dusty	M-D	M-D	dusty	Not good results
4-1299	dusty	dusty	dusty	dusty	dusty	No	No	No	dusty	dusty
2-510	M-D	M-D	M-D	M-D	M-D	M-D	dusty	dusty	dusty	M-D
2-453	No	No	No	No	No	No	No	dusty	dusty	No

The table above presents the results when applying eight folds for randomly selected dust cases to test the results of pure CBR against those of BN-CBR. The nearest neighboured 3NN was applied and the results are presented in name cells. These results indicate that CBR is able to predict the final results with an accuracy of 70%. Case 8-2557 does not have the status anticipated owing to the vote of 3NN being balanced.

Table 10. Pure CBR result in 7 folds.

ID	Result of the case	Result of 1st 3NN	Result of 2nd 3NN	Result of 3rd 3NN	Result of 4th 3NN	Result of 5th 3NN	Result of 6th 3NN	Result of 7th 3NN	Final Result
5-2546	No	No	No	No	No	dusty	No	dusty	No
3-1910	No	No	No	dusty	dusty	dusty	dusty	dusty	dusty
3-1787	M-D	M-D	M-D	dusty	dusty	M-D	dusty	dusty	dusty
7-2555	No	No	No	No	No	No	No	dusty	No
4-1023	M-D	M-D	M-D	M-D	M-D	dusty	dusty	M-D	M-D
5-255	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty	dusty
8-2557	M-D	H-D	H-D	dusty	dusty	dusty	M-D	M-D	Not good results
4-1299	dusty	dusty	dusty	dusty	dusty	No	No	No	dusty
2-510	M-D	M-D	M-D	M-D	M-D	M-D	dusty	dusty	M-D
2-453	No	No	No	No	No	No	No	dusty	No

The table above presents the results when applying seven folds for randomly selected dust cases to test the results of pure CBR against those of BN-CBR. The nearest neighboured 3NN was applied and the results are presented in name cells. These results indicate that CBR is still only able to predict the final results with an accuracy of 70%. Whereas in BN-CBR there is 20% improvement in accuracy.

Figure 37. KNN selection error.

Figure 37 above demonstrates that the number of K selection with minimum error in 3 KN was approximately 0.0155, whereas the maximum for 5 KNN is almost 0.020. The choice of K number for CBR could be determined by the size of the dataset. Given that the current study is using a dataset with in excess of 7,600 cases, the 3NN could provide the most reliable results in terms of weather status.

Table 11. Comparing algorithms results.

Algorithms		Correctly classified percentage	Incorrectly classified percentage
Pure CBR		99,18%	0.82%
Pure CHAID		87,01%	12,99%
Pure BN		76,01%	23,38%
Pure ANN		83,65%	16,35%
3NN	CHAID + CBR	97.02%	2.98%
3NN	**BN + CBR**	**98,45%**	**1,55%**
3NN	ANN + CBR	98,24%	1,76%
5NN	CHAID + CBR	96,33%	3,67%
5NN	BN + CBR	97,81%	2,19%
5NN	ANN + CBR	97,64%	2,36%

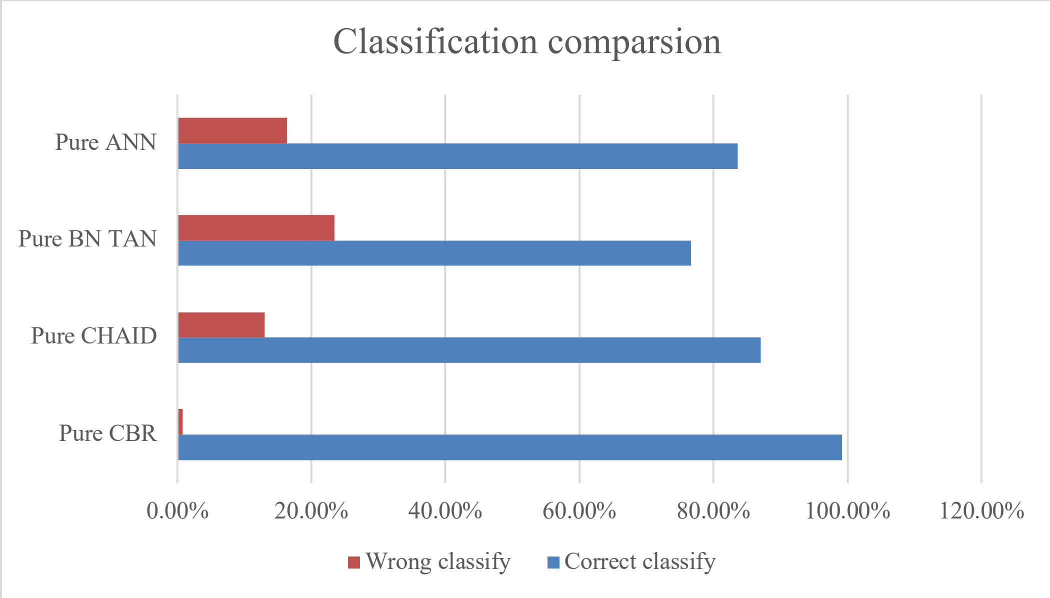


Figure 38. Classification comparison.

Figure above reveals that the BN has a low rate of classification but when combined with CBR it offers the highest rate of classification. The low rate for the pure BN classification is the result of dust being carried over long distances to the station or could have been caused by the large number of unfit dust cases in the dataset. The BN must amend the weather status of them based on the values of key attributes, when it fails to make a logical explanation.

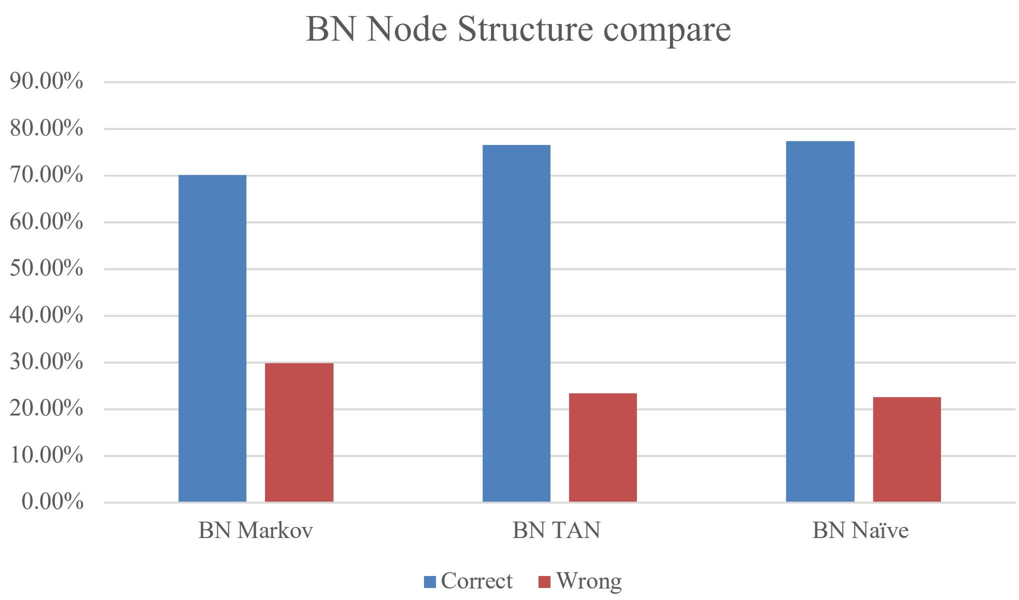


Figure 39. Comparison of BN learning algorithms.

Comparison is made between the BN learning algorithms in order to identify the one that is best able to classify dust cases with a logical explanation. Figure 38 above indicates that the naïve bays algorithm offers a higher rate of classification than the alternatives, albeit that the node structure does not explain a greater degree of correction between dust features. One of the aims of the current study is to identify an effective classification algorithm for the purposes of reliably forecasting future dust events. Therefore, the TAN is selected because it offers a better node structure for representing the relationship between inputs and the target and the correlations (both direct and indirect) between the various nodes.

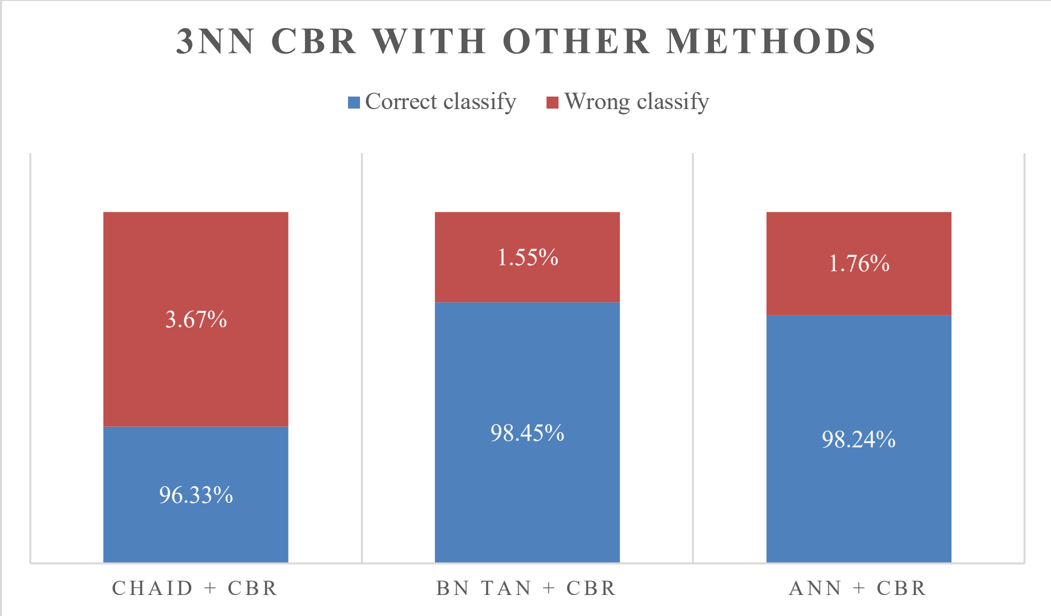


Figure 40. CBR combined with other methods.

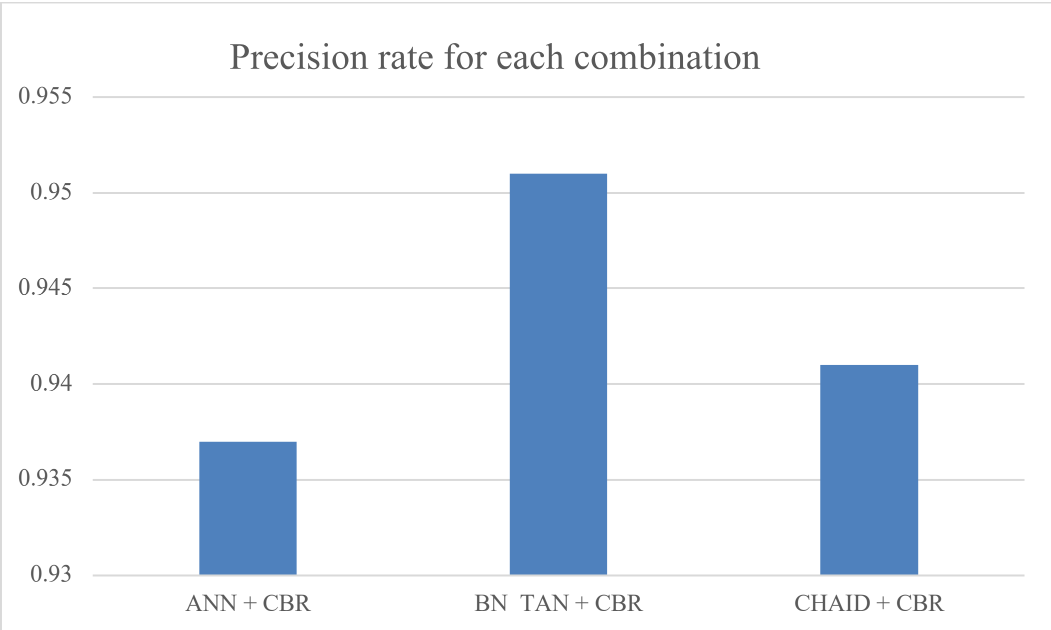


Figure 41. Precision rate of each combination.

Of all the available options, the combination of BN with CBR offers the highest rate of precision. As such, BN (TAN) combined with CBR offers the best possible forecast of future dust events compared to the available alternatives, in terms of precision rate.

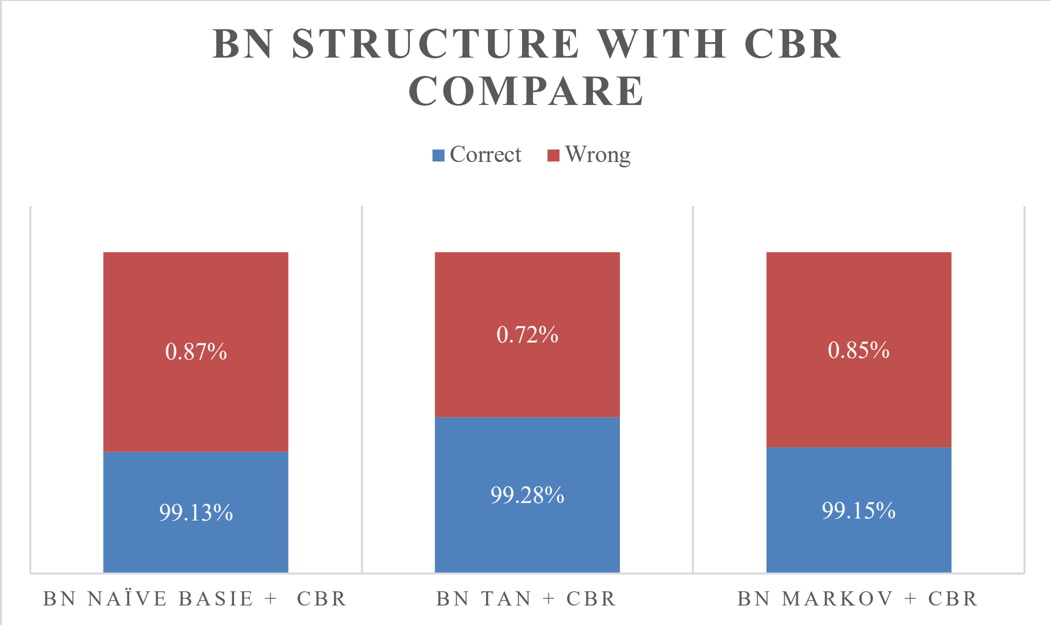


Figure 42. BN structure with CBR.

Figure above shows the benefits of combining BN, TAN and CBR. TAN is performed well because of its ability to offer logical explanations coupled with the retrieval stage with KNN in the CBR process. Therefore, it appears that the BN, TAN with CBR combination offers the most accurate classification and forecast of future dust events relative to the alternative BN learning algorithms.

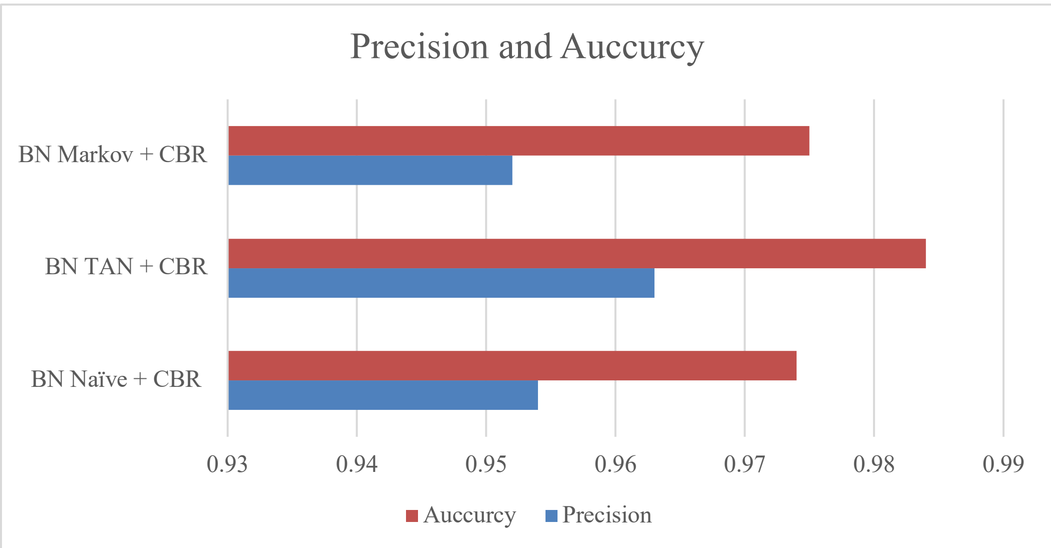


Figure 43. Precision and accuracy.

RBS and Safety Actions

Rule-based systems can be categorised into four main types according to the number of inputs and outputs: single-output-single-input, single-output-multiple-input, multiple-output single-input and multiple-output multiple-input. Any one of these may satisfy the features of association rules because the relation rules reflect the correlation between the attributes. It is possible that a relation rule will have either single or multiple rule terms in both the antecedent and consequential parts of the rule. The approach selected for the current study was single-output-multiple-input. Meanwhile, in order to identify the rules, the CHAID algorithm has been selected (see Figure 44). The CHAID algorithm is implemented using the SPSS modeller program. Compared to Quest and C.50 which produced scarce rules, CHAID has successfully generated a large number of real rules that are highly accurate.

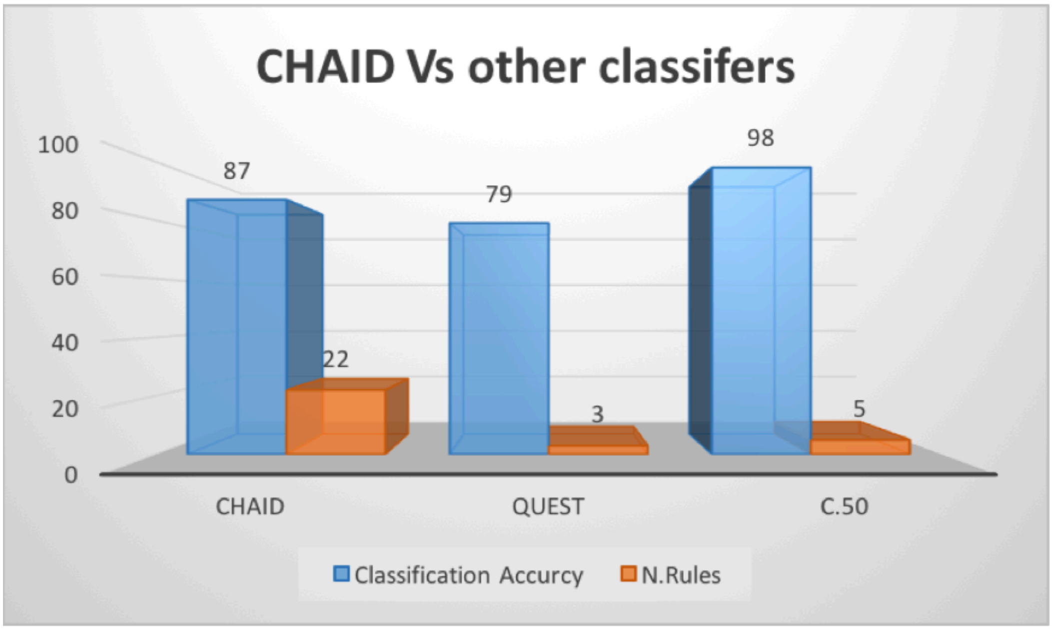


Figure 44. CHAID vs other classifiers (Al Murayziq, Kapetanakis & Petridis, 2017).

Rules

The rules generated by the CHAID algorithm were sent to the weather expert for evaluation. The evaluation came with recommendations to set suitable actions and safety standards for more effective actions. For instance, in the event that the wind speed exceeds 30km/h and the condition is medium dust, the appropriate action is to notify the health and education sectors so that they can make suitable preparations. C code requires the daily weather record or the future dust event predicted from BN-CBR. The program then runs to find an appropriate rule that can be matched to actions. .

WIND_MAX_SPEED <= 19 [Mode: dusty]
MONTH <= 4 [Mode: dusty] => dusty
MONTH > 4 and MONTH <= 5 [Mode: dusty] => dusty
MONTH > 5 [Mode: dusty] => dusty
WIND_MAX_SPEED > 19 and WIND_MAX_SPEED <= 22 [Mode: dusty]
PRESSURE_MIN_STATION_LEVEL <= 937.600 [Mode: dusty] => dusty
PRESSURE_MIN_STATION_LEVEL > 937.600 and PRESSURE_MIN_STATION_LEVEL <= 938.400 [Mode: m-Dust] => m-Dust
PRESSURE_MIN_STATION_LEVEL > 938.400 and PRESSURE_MIN_STATION_LEVEL <= 939.800 [Mode: h-Dust] => h-Dust
PRESSURE_MIN_STATION_LEVEL > 939.800 [Mode: dusty] => dusty
WIND_MAX_SPEED > 22 and WIND_MAX_SPEED <= 25 [Mode: dusty]
PRESSURE_MAX_STATION_LEVEL <= 937.200 [Mode: h-Dust] => h-Dust

Figure 45. Sample of rules generated by CHAID algorithm(Al Murayziq, Kapetanakis & Petridis, 2017).

All of the generated rules are based on the key dust characteristics which determine the importance of them before using the same algorithm based on the SPSS Modeller. The above figures show the original rules that are generated but after closer inspection it seems there is a need for improvement because more than one rule contains confusion cases in order to be able to write them in C# code using if statements. For example, the first rule contains unnecessary sub rules which could make the rule difficult to understand. It would be easier to read and understand if it was as follows:

Before:

WIND_MAX_SPEED <= 19 [Mode: dusty]

MONTH <= 4 [Mode: dusty] => dusty

MONTH > 4 and MONTH <= 5 [Mode: dusty] => dusty

MONTH > 5 [Mode: dusty] => dusty

After:

WIND_MAX_SPEED <= 19 [Mode: dusty]

MONTH <= 4 and MONTH > 5 [Mode: dusty] => dusty

In the above rule, the if statement could be written in a more simple way while still maintaining the same meaning. By measuring the confidence of the rules it is possible to determine how effective they are as well as the benefits that can be achieved from building a safety action on it. This is important because poor rules will result in false alarms. In addition to the confidence of the rules, the weather expert advocated the logic of the rules by changing them as necessary.

Weather Experts Evaluating Rules

Once the rules had been generalised and written in evaluation form, they were sent to Saudi weather experts. Dr almisind and his colleagues endorsed the rules and noted that the rules contain special weather conditions so as to better interpret unusual dust events. The following table presents a section of the evaluation form. The team of weather experts offered their opinions on each of the dust categories, reflecting on what they saw and thought must accrue to create each type of dust event. This process could significantly improve the rules confidence.

Table 12: Weather expert's comments on dust cases.

Rules	Conditions	1st Expert	2nd Expert	3rd Expert
Month: March to July	Dusty	Dust events are common in these months and a wind speed of just 15km/h could result in dusty conditions	Dusty conditions could occur in this period and wind speed is the factor that has the greatest effect under this rule	A combination of wind, low pressure and temperature are required
	Medium dust	Medium dust is common in these months but this relies on other dust factors	There must be no humidity, a wind speed in excess of 22km/h and a high temperature	Medium dust events require wind speed and low pressure
	Heavy dust	Low pressure and windy conditions are required	Winds in excess of 30km/h, pressure below 935 and high temperatures	The most important factors are high wind speed and very low pressure
Wind speed <= 22 and pressure => 950.	Dusty	While there is no need for low pressure, there must be sufficient wind, a high temperature and low humidity	In desert terrain, wind alone can create dusty conditions	A wind of this speed could cause dusty weather, especially in summer when humidity is low
	Medium dust	These conditions are not ideal for medium dust because a higher wind speed would be required	Not ideal conditions for medium dust	Not ideal conditions for medium dust
	Heavy dust	Not ideal conditions for heavy dust	A higher wind speed would be required	The wind speed would need to increase; therefore, not ideal conditions
Wind Speed > 22 and pressure < 946	Dusty	This could cause dusty conditions even if there is high humidity and low temperatures	Dusty weather could result even if it has recently rained	Could result in dusty conditions if humidity is high and the surface is moist
	Medium dust	If humidity is not high, a wind speed of between 22 and 28km/h could cause medium dust	This wind speed and pressure could generate medium dust	In low humidity, wind speeds in excess of 27km/h could cause medium dust events even in the absence of low pressure
	Heavy dust	The ideal conditions are low pressure, low humidity and wind speeds in excess of 30km/h	Heavy dust requires dry ground conditions, low pressure and wind speeds in excess of 29km/h	If the surface is dry, high winds with pressure of less than 937 could result in heavy dust

Rules Confidence

Actions

In order to minimise the harm that dust events cause to human health and agricultural production, a set of actions must be agreed. When conducting a search for the literature review, there did not appear to be a set of actions that weather experts had previously recommended that could be applied as standard safety actions. Therefore, when a region experiences extreme weather conditions, actions to protect the general public are only agreed at that time. Advice is sought from weather experts and other leading professionals to devise a plan of action for the anticipated conditions be that heavy rain, fog or dust events. Therefore, it would potentially be beneficial to devise a safety standard that could be deployed for a given weather condition. The current study seeks to forecast dust events and provide supporting advice regarding actions that should be taken based on input from weather experts. All actions have been evaluated for dust events by weather experts at the Saudi meteorological department. Seeking professional advice lends support to the advice offered, thereby helping to enhance its credibility. These actions can be adjusted when a dust event occurs based on the input from weather experts.

Weather Expert's Comments on Dust Type Actions

Three experts from the meteorology department offered their opinions on the actions that should be taken when facing different dust events. These opinions can then be used to develop initial actions to help mitigate the adverse effects of dust events. These actions should eb agreed for several reasons. First, to help apply the rules based system and support decisions. Secondly, to devise a suitable safety standard for dust events. Over time, this standard could be revisited to make changes for each individual dust category if necessary.

Table 13: Weather expert's comments on safety actions.

	Conditions	1st Expert	2nd Expert	3rd Expert
Actions	Dusty	No particular actions are needed in dusty weather	In dusty weather the public should be aware of the need to take action if the dust becomes more intense	Most dusty weather occurs in the upper atmosphere so it is only necessary to remain aware and watch the latest forecasts
	Medium dust	The public should prepare for changes in outdoor activities	Precautionary preparations should be made and the public should be informed via the media	People should be told to stay indoors where possible because the main effect of medium dust is on traffic and human health
	Heavy dust.	People need to stay aware and be aware of what actions will help them	Heavy dust adversely affects human health so steps should be taken including closing schools	When faced with heavy dust, the public should stay indoors and take any other actions they see fit

The initial actions that should be taken when facing the different dust events are set out below:

1. Dusty: Be aware

   - Teach the general public that dusty conditions can have adverse effects on their vision and health
   - Be aware of the latest weather forecasts
   - Visibility can be greatly reduced, especially at night time

2. Medium dust storm: Get prepared

   - Be aware of the latest weather forecasts
   - Contact schools and health centres to inform them of the latest weather conditions
   - Take steps to prepare for anticipated traffic congestion
   - Use the media to tell the general public what steps they should be taking
   - Firmly close all windows and doors
   - If it is necessary to venture outside, be sure to use a dust mask

3. Heavy dust storm: Take action

- Issue advice to avoid going outdoors in affected regions
- Firmly close all windows and doors
- Dampen cloths and lay them under doors to prevent dust entering buildings
- If it is necessary to venture outside, be sure to use a dust mask
- Contact schools to inform them of the latest weather conditions so that they can close if necessary
- Contact health centres to inform them of the latest weather conditions so that they can prepare for an increase in people attending the emergency ward and increased demand for services used to treat respiratory problems
- Use the media to inform the public about what is happening

RBS in C# Prototype

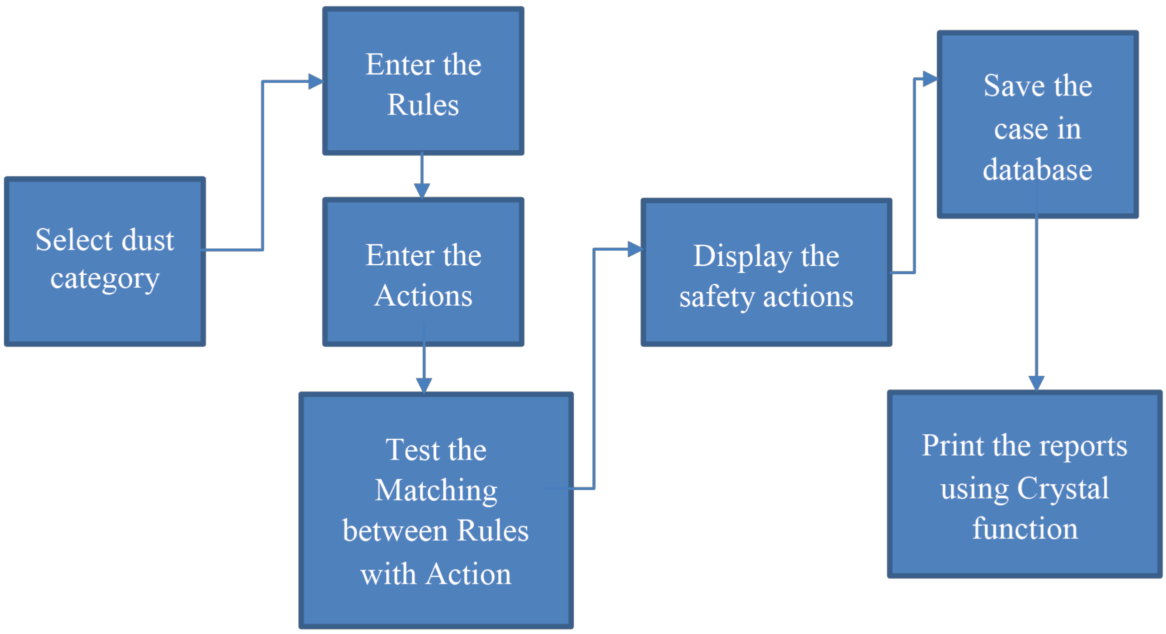


Figure 46. Proposed RBS prototype.

The most effective way to match the rules and actions is to develop a prototype to better understand the role played by the RBS. Rather than relying on conventional rule based tools such as CLIPS, the decision has been taken to program a prototype because programming language is a preferred way to understand this weather condition to the meteorological department, also makes the user experience more efficient when presenting results using the UI. Furthermore, a prototype means that users deal with rules as inputs and actions as outputs. Although the process of writing the code is laborious and time-consuming (especially for IF-then statements when there are many rules), this effort is worthwhile because of the efficiency benefits it affords when displaying the role of RBS. Using CLIPS to enter rules and actions and run the system is relatively simple and is not time-consuming and it arrives at the same results as when using programming language. In addition, CLIPS presents the results as facts. However, CLIPS presents the results in a very basic format and this is why C# is preferred because the results will be identical to the expert system tools. Importantly, CLIPS presents the results in an efficient manner. Below is an illustration of how the C# code is able to use an IF-then statement to match the rules and actions.

Figure 47. The main window of the RBS prototype.

There are three options on the main page. The first is concerned with entering the rules from the CHAID algorithm after being refined. Using the same option it is possible to test the rules by matching them with actions and searching dust events contained within the database. The second option allows you to add actions that have been derived from the opinions of weather experts for each of the different dust events. Finally, the third option concerns the prototype.

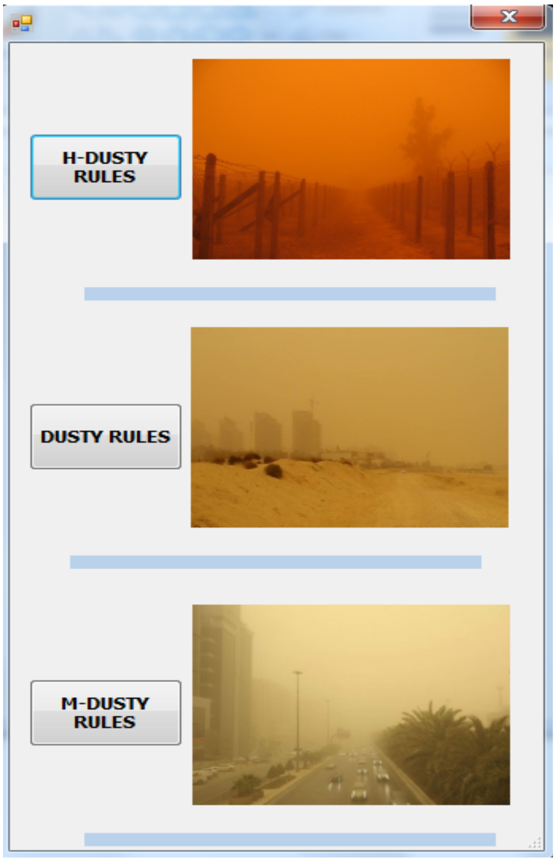


Figure 48. To fill the rules in a different category.

The main rules page offers three options depending on the type of dust. This distinction helps to ensure that the correct rules are added for each type of dust. In addition, it makes the process of amending or deleting entries easier. The database has been divided into different classes and there are three rules windows. In each window, rules can be entered and saved under various labels to reflect the relevant category.

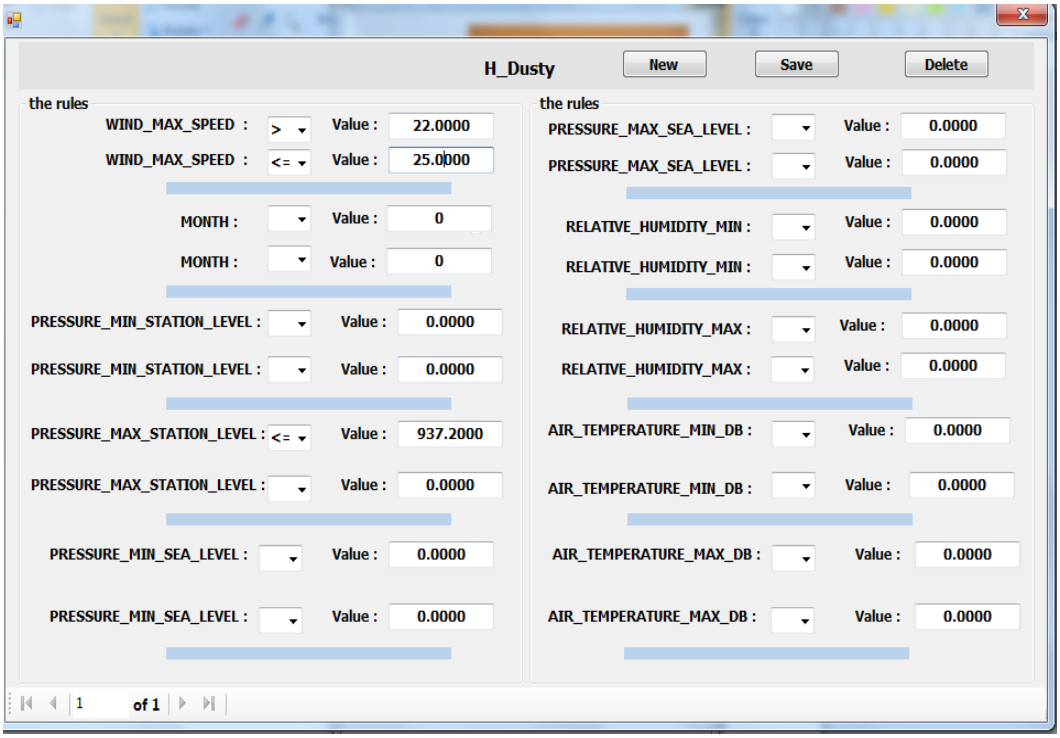


Figure 49. Window to fill heavy dust rules.

This window enables the user to add the rules for heavy dust. The above figure lists the key attributes which are based on CHAID outcomes. It may be that the attributes in the rules have two values; e.g. Pressure_min_station > 934 and < 940. The window above provides no option to add further dust features and the research relies on the characteristics identified by the weather expert and the various algorithms as being the features that have greatest impact on dust events. There are three functions in the window: 'New' to add rules, 'Save' to save the values that have been entered and 'Delete' to delete rules from the page indicator.

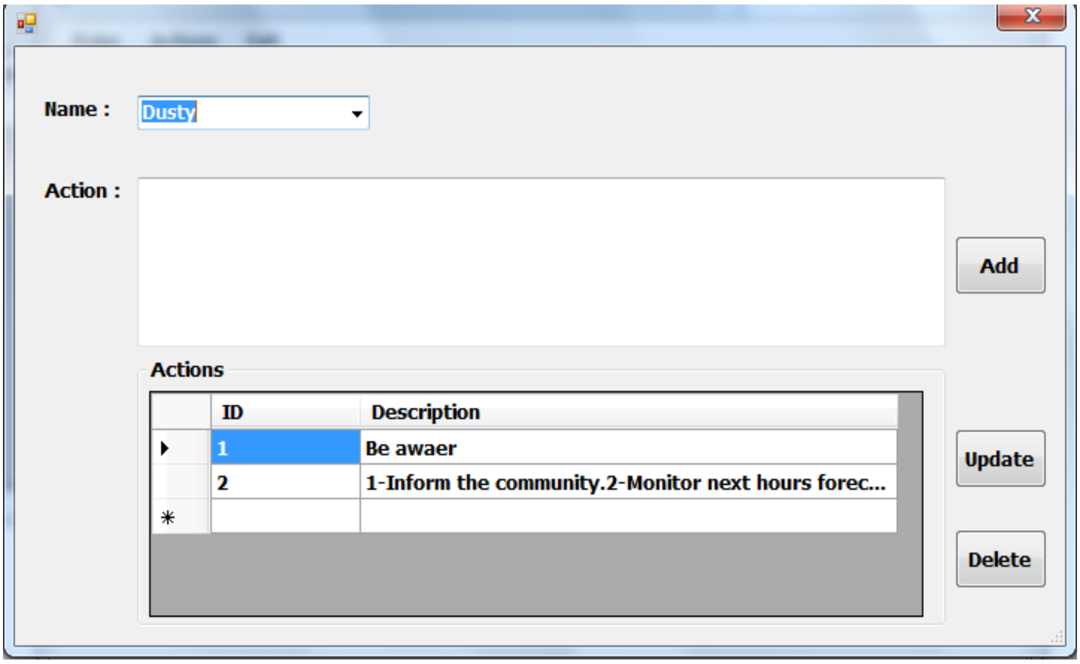


Figure 50. Window to add actions for each dust category.

This window enables the user to add actions and instructions to help mitigate the adverse effects of dust events. Reference is made to each of the different dust categories covered by the current study and a drop-down menu allows the user to select a type of dust and fill in appropriate actions. In addition, a code can be added to represent a set of actions if they are too numerous. The actions tally perfectly with the rules that have previously been entered. Each of the actions can be modified or deleted at a later date.

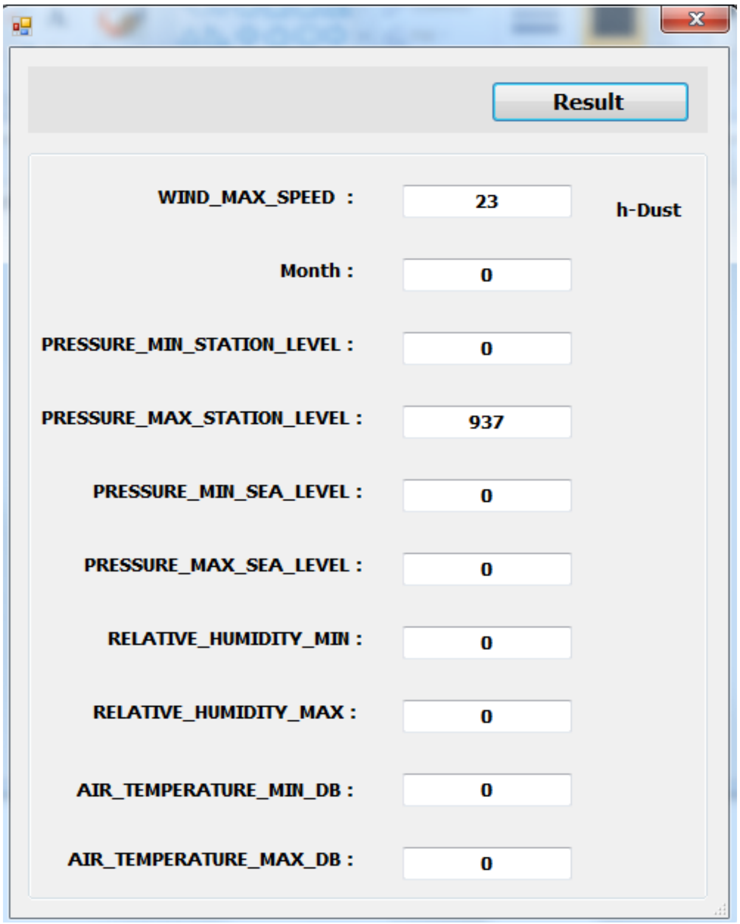


Figure 51. Test window to test the matching between rules and action.

The weather reading must be entered as inputs in the testing window. Having entered these values, the IF-statement searches for rules with the same reading range so that the rules can be executed. It is not necessary to fill all fields based on the critical weather reading. For example, if there is low pressure and high wind speed, a high dust category may be selected by the IF functions if the inputs lie within the range of the stored rules. Pressing the results button causes the proper action to be displayed and the new dust cases can be saved with the weather reading and action.

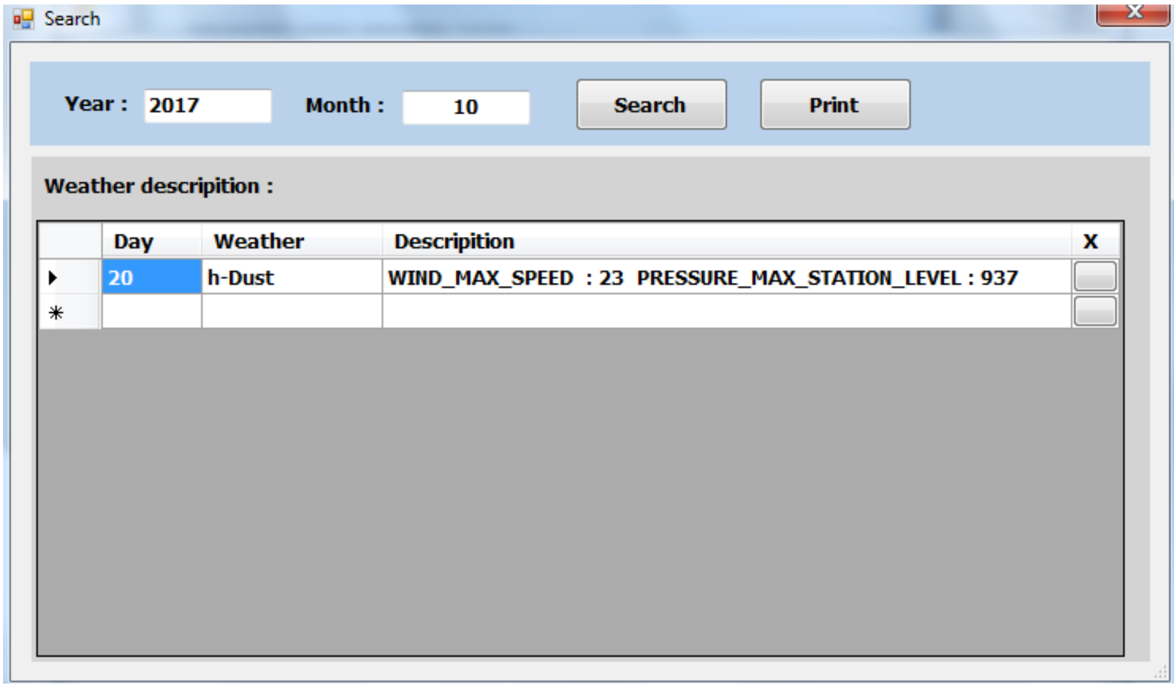


Figure 52. Search window to display the saved dust cases.

All of the saved dust events are displayed in the search window where they are presented in date order. Selecting a month and year reveals all dust events in that period. It is also in this window that certain events can be deleted or printed.

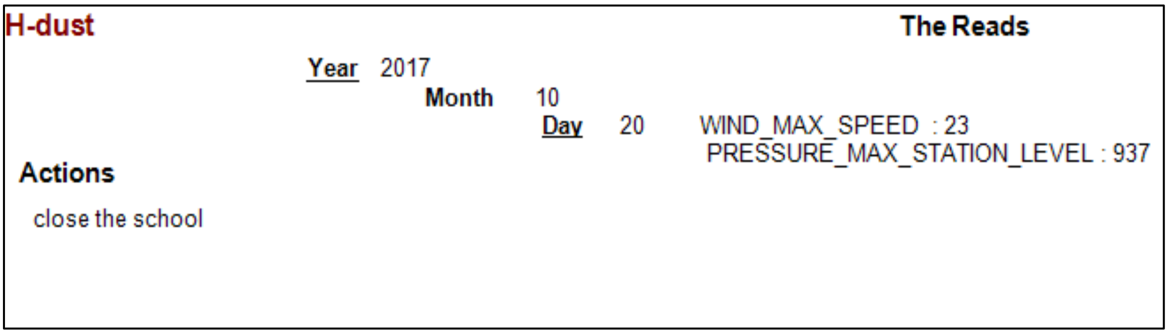


Figure 53. Dust cases extracted as report using Crystal function.

The decision to write RBS with a programming language enables a dust archive to be amassed that combines weather readings with appropriate actions. This facility will enable weather experts to not only better understand dust events but also to evaluate actions in a timeline. Data including the date, dust category, weather reading and actions are compiled and the Crystal report allows these details to be extracted in the form of an Excel spreadsheet, thereby enabling researchers to utilise the database for study purposes. The Crystal function was designed using the same visual studio used to program RBS with C# language.

Chapter Five

Validation and Evaluation

Validation

Dataset and Reality

One of the main factors governing the success or failure of a research study is the choice of data. In order for the current study to arrive at relevant conclusions, the dataset must satisfy a number of criteria. First, the data must span a period of at least fifteen years and consist of recent data. Even then, a fifteen year period is relatively short in meteorological terms and given the geology of the region because sustained periods of drought are associated with dust events. Saudi Arabia comprises vast areas of desert and areas affected by drought and these arid areas have expanded in recent years owing to the effects of human activity, including climate change. Secondly, it is imperative that the dataset contains all of the key dust characteristics that have been identified including temperature, pressure, wind speed and humidity. Not all datasets include the full range of variables and some only include mean values for each of the variables. Relying on data of this nature would significantly affect the relative importance of the variables and is highly likely to alter the order of the key features of dust events. For example, it is necessary to record the maximum wind speed and not the mean figure because as the wind speed accelerates (and in combination with other factors), dust storms are more likely to occur. However, if only mean wind speeds are recorded, this will have the effect of diminishing the importance of the wind speed variable. Thirdly, in order for the dataset to be effective and generate accurate results, it is necessary to include noisy cases (i.e. record all dust readings). However, in the final weather status there are no dust storms, and this may be because the recorded case is affected by a preceding condition such as rain which makes it considerably more difficult for the wind to carry wet or muddy soil. Alternatively, the winds may have been travelling from a direction that is free from drought surfaces. It may also be that the last possible weather status was not recorded which happens relatively frequently when using old data. Finally, it is important that there must not be any missing vales in the dataset; especially for the key dust attributes. This is because the dataset must contain the true weather readings and not merely assumptions that have been added by the researcher. In order to truly understand the condition and the case in dust events, true records must be used.

BN Effective to Project

In order to test the validity of BN it is necessary to satisfy all of the requirements for the capability of BN. This requires all data inputs to be in numeric form and ordered so they can be read using the Weka tool. For example, the values for the surface attributes are divided into four categories to indicate the type of surface. Therefore, these categories must be converted from string to numeric form. In order for the model to be optimised, the target must ensure that the final status is recorded. Furthermore, the BN algorithm works best when there is a degree of similarity between the cases and predictions or classifications can be made according to probability. This work, the BN is used to classify uncertain cases. This process requires knowledge of the condition of any uncertain cases and these will be classified by the BN algorithm into the various dust categories. In order to avoid errors at this stage, cases without a weather status that have no recorded range of dust scale are removed from the dataset. Doing so ensures that the BN will not produce erroneous results. The datasets provided do not have a direct weather status assigned to them but there is a weather code for each case which does indicate the weather status. This can be used in the BN to achieve reliable results. However, first it is necessary to assign the dust events into the relevant categories so that the algorithm can gain a full appreciation of the different categories. For the purposes of the current study, three dust categories are used.

CBR Cycle and Similarity Between the Cases

CBR is employed in this work to forecast future dust events. In order to achieve accurate results, it is necessary to prepare all of the factors used in CBR. The first stage in this process is to determine the most effective attributes for the problem domain. Isolating the key components of the available dataset makes it possible to explain, reason and understand the cases. Once these key features have been determined, it is necessary to check the similarity between cases in order to ensure that the data are appropriate for use with CBR. CBR can only be used if there is a high degree of similarity. If this is not the case, an alternative method would be required. The primary purpose of using CBR is to match similar cases in order to solve problems. Therefore, if there is not a high degree of similarity between the dust cases, it would not be appropriate to apply CBR. Furthermore, similarity measurement equations consider the weight of the various attributes to improve the accuracy of the results. Certain attributes will play a more important role than others and this must be reflected in the choice of weights. This can be achieved using a weighted similarity measurement such as the weighted Euclidean distance. For the CBR to operate smoothly, it must follow the

CBR cycle and care should be taken to ensure with each step starting with obtaining information for the database, establishing similarity, re-using old solutions from similar cases to new cases, and then revising the matching and ensuring that the solution is capable of effectively solving the problem at hand. Finally, any modifications to the database that are deemed necessary should be made.

Having established the degree of similarity between the cases and matching them with new cases, the accuracy of the results can be enhanced by applying the nearest neighbour (NN) technique. This is able to improve the accuracy of the results by using the vote for the result of certain cases. In the current research study, 3NN has been employed to arrive at the best possible prediction (as mentioned in K selection error). It is necessary to determine the k because a large k means that it is less sensitive to noise (especially class noise), better probability estimates for discrete classes and large-scale training sets help to enhance the value of k. A small k is better able to capture the fine structure of the space problem and this may be necessary if the training set is relatively small. A suitable balance must be achieved between small k and large k as the training set approaches infinity with k growing large and KNN becoming Bayes optimal.

Rules and Actions

Using a rule-based system will enable weather experts to take appropriate actions by directing them according to the rules contained within the database. CHAID algorithms are used to extract rules in this work but it may be necessary to amend some of the rules to make them more effective and deliver the desired outcomes. For this to be possible it is necessary to ensure that each group represents certain cases or categories by revising the rules and checking that they are comprehensive for all type of dust. The authenticity of the rules must also be checked by measuring the confidence for each rule according to the strength of the relationship with each particular scenario. One possibility is to ask the weather experts to evaluate the rules based on their extensive knowledge of this field and their comprehension of the available data. Doing so could well help to increase the confidence rate of some of the rules. It is important to note that all of the rules applied in this work were issued by the database for the cities of Riyadh and Dammam which were chosen due to the specific nature of the terrain in these locations and the unique climate of the Gulf sea. Furthermore, if the database were to change to alternative cities, they may have different climates and require the rules to be amended.

The efficacy of the actions must also be confirmed for the anticipated dust cases. Owing to the lack of standard actions in response to different dust events and the similar absence of regulations governing such matters, it was deemed necessary to create a set of safety actions that can be implemented when different dust cases occur. These actions will vary depending on the type of dust event, with all actions being evaluated by the Saudi meteorological authority.

Evaluations

The entire approach will be subjected to tests in order to confirm that it is consistent and yields high quality results that closely mirror reality. The current section is therefore, of considerable importance because it includes an evaluation of the results and comparison of the obtained results with reality. This involves comparing the results in the list and the warnings issued by the meteorological department. Furthermore, in an attempt to ensure the most accurate results possible are achieved in this work, the results will be compared with current models which are used by the majority of meteorological authorities. This involves choosing a number of future cases at random without making a note of their weather status and then using the BN-CBR classification results and the initial predictions that were recorded. These predictions will then be compared with satellite images. In order to produce accurate results and assess the methodology fairly, a large number of future cases must be tested in order to gain an appreciation of the accuracy of the average prediction.

For the purposes of the current study, tests were conducted for twenty future cases in different months. It was important to choose cases from different months owing to the considerable differences in wind speeds over the course of the year. All of the cases were taken from the Riyadh dataset over the period April-September 2017. Once the prediction has been noted, the RBS is used to set a group of actions to help mitigate the effects of the dust event. All of the actions will have previously been evaluated by weather experts but the rule based system will be tested using real cases to confirm how effective the actions are.

Cross Validation (CV) and Other Evaluation Equations

K-fold cross-validation is used to evaluate the predictions. This involves dividing the data into k subsets, all of which are the same size. The models are built k times and one of the subsets is omitted each time. This is used as the test set. If k is the same as the sample size, this is referred to as leave-one-out. A single estimation can be arrived at by averaging the k results from the folds. This is beneficial because it ensures that all observations are used for training purposes and also for validation. Moreover, each of the observations is used a single time for validation. Meanwhile, stratified k-fold cross-validation is often used for the purposes of classification. This involves folds being selected in a way that each fold comprises approximately the same proportions of class labels.

- MSE gives an average of the squares of the errors or deviations (difference between the estimator and what has been estimated)
- RMSE is the square root of the arithmetic mean of the squares of a set of numbers (a measure of imperfection of the fit of the estimator to the data)

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=0}^{n} (\boldsymbol{Pi} - \boldsymbol{Oi})^2}{\boldsymbol{n}}} \tag{7}$$

- MAE measures the accuracy of the forecasts relative to the eventual outcomes
- R2 is the proportion of the variance in the dependent variable that is predicted using the independent variable

$$\boldsymbol{R2} = \mathbf{1} - \frac{(\mathbf{1} - \boldsymbol{R}^2)(\boldsymbol{N} - \mathbf{1})}{\boldsymbol{N} - \boldsymbol{P} - \mathbf{1}} \tag{8}$$

Various evaluation methods have been employed to the dataset for the purposes of comparing the results and establishing how accurate the outcomes are. One of the methods used is leave-one-out (LOOCV) which is the approximate function trained on all data with the exception of a single point and it is on this point that a prediction is based. The average error is estimated and this forms the basis of efforts to evaluate the model. As can be seen below, the LOOCV does not indicate a significant variation in the accuracy of predictions, possibly due to the nature of the problem domain. For instance, weather forecast are usually highly accurate and, therefore, applying alternative evaluation methods is unlikely to offer any significant improvement. In addition, a random split between the test and training makes it possible to appraise the results. When the dataset if split 30, 70 or 10, 90, there is no significant improvement. A split of 40, 60 wasn’t applied to the data owing to the fact that the split percentage of the CV is very close. Therefore, it is deemed that the most appropriate evaluation methodology when arriving at the prediction result is the CV.

Test & Score

Settings

Sampling type: Stratified 10-fold Cross validation
Target class: Average over classes

Scores

Method	AUC	CA	F1	Precision	Recall
kNN	0.998	0.979	0.987	0.973	1.000
Neural Network	0.999	0.979	0.989	0.979	1.000
Random Forest	1.000	0.997	0.998	0.997	1.000

Figure 54. Cross validation outcomes.

Test & Score

Settings

Sampling type: Shuffle split, 10 random samples with 90% data
Target class: Average over classes

Scores

Method	AUC	CA	F1	Precision	Recall
kNN	0.998	0.978	0.986	0.973	1.000
Neural Network	0.999	0.976	0.987	0.975	1.000
Random Forest	1.000	0.997	1.000	1.000	1.000

Figure 55. 10, 90 random split evaluation.

Test & Score

Settings

Sampling type: Shuffle split, 10 random samples with 30% data
Target class: Average over classes

Scores

Method	AUC	CA	F1	Precision	Recall
kNN	0.997	0.951	0.970	0.942	1.000
Neural Network	0.998	0.956	0.972	0.957	0.987
Random Forest	1.000	0.991	0.997	0.993	1.000

Figure 56. 30, 70 random split evaluation.

Test & Score

Settings

Sampling type: Leave one out
Target class: Average over classes

Scores

Method	AUC	CA	F1	Precision	Recall
kNN	0.998	0.980	0.987	0.974	1.000
Neural Network	0.999	0.979	0.991	0.983	1.000
Random Forest	1.000	0.998	0.999	0.999	1.000

Figure 57. Leave-one- out evaluation.

Hybrid Intelligence System in Real World Prediction

Dust predictions are presented, future weather records have been used to conduct an experiment that confirms the effectiveness of the chosen methodology. Based on these results, it has been possible to correct some of the errors that have become apparent. Testing the selected weather cases reveals the true situation and the forecasts arrived at using hybrid techniques will be presented so that comparison can be made with the model that the Saudi meteorology department currently uses. The first step is to take the prediction generated by the BN and support this with the prediction derived from CBR. Clearly, this provides two separate predictions that have been generated using different methodologies. Whereas the BN classifies new cases on the basis of probability, CBR arrives at forecasts based on how closely the current case mirrors a sample of historical cases. The evaluation is initially conducted using only the BN but subsequently the CBR prediction will also be incorporated. Finally, the RBS will suggest suitable course of action to help mitigate the effects of the dust event. Weather experts will then appraise the predictions and the suitability of the suggested actions in order to determine whether the chosen methodology could be adopted for use in the real world. The model used to compare the results with that of the model is called the Automated Early Warning System (AEWS) and it is this system that is used to provide early warning alerts of the weather phenomena such as heavy rain, dust events and hurricanes that are issued by the General Authority for Meteorology and Environmental Protection in Saudi Arabia. The AEWS produces its alerts based on satellite images and the resulting alarms are issued in one of four categories. 'Notice' suggests that close attention needs to be paid to the emerging weather data so that any change in weather status is observed in a timely manner. ‘Alert’ indicates that there is a possibility of an area being affected by adverse weather. ‘Advanced Alert’ signals that there could be adverse effects resulting from the latest weather developments and caution is required. Finally, ‘Warning’ points to severe meteorological phenomena and, therefore, full precautions must be taken including notifying the civil defence and issuing guideline instructions to the parties that are likely to be affected. For the purposes of the current study, the last three categories are compared with the RBS outcomes in order to appraise the process associated with the rules and actions.

Results of Hybrid Intelligence System Compared with Current Model

The prediction results for the BN-CBR are presented in this section and subsequently compared with the forecasts based on satellite images. The first results are for the BN classification of new dust cases and the initial prediction attempts. This is followed by the CBR results using data for all new dust events in order to arrive at a final prediction.

Table 14. Comparing the approach with current model results.

Date	BN classification	CBR prediction	AEWS prediction	Real weather status
11/4/2017	h-dust	h-dust	Advanced alert	Dust storm
25/4/2017	h-dust	h-dust	Advanced alert	Dust storm
9/6/2017	h-dust	h-dust	Advanced alert	Dust storm
3/7/2017	h-dust	h-dust	Advanced alert	Dust storm
18/5/2017	h-dust	h-dust	Advanced alert	Dust storm

The above table confirms the BN classification tallies with the CBR predictions and also with the results of the AEWS model. As previously mentioned, the BN algorithm assigns cases to a dust category on the basis of probability, whereas for the CBR prediction the BN results were imported with a prediction margin and historical dust events were used to arrive at the closest possible match. The BN prediction margin is an input in the CBR process and helps to improve the accuracy by matching the current scenario with historical episodes. The following table provides an example of how the BN classifies new cases and how the CBR is able to forecast the appropriate weather status for a given dust event.

Table 15. New recorded cases:

Date	**25/4/2017**
WIND_MAX_SPEED	55
PRESSURE_MAX_STATION_LEVEL	943.6
PRESSURE_MAX_SEA_LEVEL	1009.8
AIR_TEMPERATURE_MAX_DB	36
RELATIVE_HUMIDITY_MAX	34
PRESSURE_MIN_STATION_LEVEL	936
PRESSURE_MIN_SEA_LEVEL	1002.3
AIR_TEMPERATURE_MIN_DB	22.8
RELATIVE_HUMIDITY_MIN	7
RAINFALL_TOTAL	0

Table 16. Case after BN classification:

Date	**25/4/2017**
WIND_MAX_SPEED	55
PRESSURE_MAX_STATION_LEVEL	943.6
PRESSURE_MAX_SEA_LEVEL	1009.8
AIR_TEMPERATURE_MAX_DB	36
RELATIVE_HUMIDITY_MAX	34
PRESSURE_MIN_STATION_LEVEL	936
PRESSURE_MIN_SEA_LEVEL	1002.3
AIR_TEMPERATURE_MIN_DB	22.8
RELATIVE_HUMIDITY_MIN	7
RAINFALL_TOTAL	0
Prediction margin	0.999988
BN Classification category	h-Dust

The BN classification utilises historical cases to build a model and assign current cases to an appropriate category. As previously indicated, the prediction margin closely matches the h-dust category.

Table 17. CBR prediction:

Date	**25/4/2017**	
WIND_MAX_SPEED	55	
PRESSURE_MAX_STATION_LEVEL	943.6	
PRESSURE_MAX_SEA_LEVEL	1009.8	
AIR_TEMPERATURE_MAX_DB	36	
RELATIVE_HUMIDITY_MAX	34	
PRESSURE_MIN_STATION_LEVEL	936	
PRESSURE_MIN_SEA_LEVEL	1002.3	
AIR_TEMPERATURE_MIN_DB	22.8	
RELATIVE_HUMIDITY_MIN	7	
RAINFALL_TOTAL	0	
Prediction margin	0.999988	
Result of 3 KNN	h-Dust	0.9
Result of 5 KNN	h-Dust	0.98
Result of 10 KNN	m-dust	0.89
Final Prediction	H-dust	

New cases are taken from the BN outcomes dataset and the CBR retrieves historical dust cases that best match the new case. It does so by measuring the distance between the current case and the historical cases and utilises 3 KNN. It is anticipated that the predictions resulting from the BN-CBR methodology will closely mirror what actually occurs.

Table 18. New recorded cases:

Date	**26/4/2018**
WIND_MAX_SPEED	45
PRESSURE_MAX_STATION_LEVEL	941.5
PRESSURE_MAX_SEA_LEVEL	1001.8
AIR_TEMPERATURE_MAX_DB	37
RELATIVE_HUMIDITY_MAX	20
PRESSURE_MIN_STATION_LEVEL	933.5
PRESSURE_MIN_SEA_LEVEL	1000.3
AIR_TEMPERATURE_MIN_DB	20.3
RELATIVE_HUMIDITY_MIN	9
RAINFALL_TOTAL	4

Table 19. Case after BN classification:

Date	**26/4/2018**
WIND_MAX_SPEED	45
PRESSURE_MAX_STATION_LEVEL	941.5
PRESSURE_MAX_SEA_LEVEL	1001.8
AIR_TEMPERATURE_MAX_DB	37
RELATIVE_HUMIDITY_MAX	20
PRESSURE_MIN_STATION_LEVEL	933.5
PRESSURE_MIN_SEA_LEVEL	1000.3
AIR_TEMPERATURE_MIN_DB	20.3
RELATIVE_HUMIDITY_MIN	9
RAINFALL_TOTAL	4
Prediction margin	0.999
BN Classification category	h-Dust

Table 20. CBR prediction:

Date	**25/4/2017**	
WIND_MAX_SPEED	45	
PRESSURE_MAX_STATION_LEVEL	941.5	
PRESSURE_MAX_SEA_LEVEL	1001.8	
AIR_TEMPERATURE_MAX_DB	37	
RELATIVE_HUMIDITY_MAX	20	
PRESSURE_MIN_STATION_LEVEL	933.5	
PRESSURE_MIN_SEA_LEVEL	1000.3	
AIR_TEMPERATURE_MIN_DB	20.3	
RELATIVE_HUMIDITY_MIN	9	
RAINFALL_TOTAL	4	
Prediction margin	0.999	
Result of 3 KNN	h-Dust	0.9
Result of 5 KNN	h-Dust	0.98
Result of 10 KNN	h-dust	0.89
Final Prediction	H-dust	

Date	BN classification	BN-CBR prediction	AEWS prediction	Real weather status
26/4/2018	h-dust	**h-dust**	**Late Advanced alert**	Dust storm

The previous example demonstrates the need to develop an accurate forecasting methodology because an effective early warning system (even without providing an accurate forecast of the time at which the dust event will strike) will help to avoid the negative consequences typically associated with heavy dust events such as those that occurred in Qassim and Riyadh on 26 April 2018. On that occasion, the AEWS system which relies on interpreting satellite images was only able to provide a warning of the impending dust event at very short notice, thereby giving little opportunity to take steps to mitigate the effects of the storm.

Figure 58. Riyadh dust storm on 26 April 2018. *(K. Mohammed [Photographer], Riyadh, 2018).*

Figure 59. Riyadh and Al Qassem dust storm on 26 April 2018. *(J. Nasser [Photographer], Al Qassem, 2018).*

Clear advice was not provided to the health and education sectors and, therefore, they did not know whether it would be necessary to send students home early from classes or for the hospitals to prepare medical staff to receive an influx of patients. Had warnings been issued in plenty of time, this confusion could have been avoided. This particular dust storm occurred by Shamal which is a particularly hot and arid dusty strong wind from a northerly or north-westerly direction that affects Iraq, Iran and the Arabian Peninsula. Strong winds are usually constant during the months of June and July, albeit that wind speeds rarely exceed 50km/h (30mph). However, this is sufficient to generate great dust storms, especially during July when it is common for Baghdad to experience at least five storms. The Shamal is caused by air flowing towards a large area of low pressure over Pakistan.

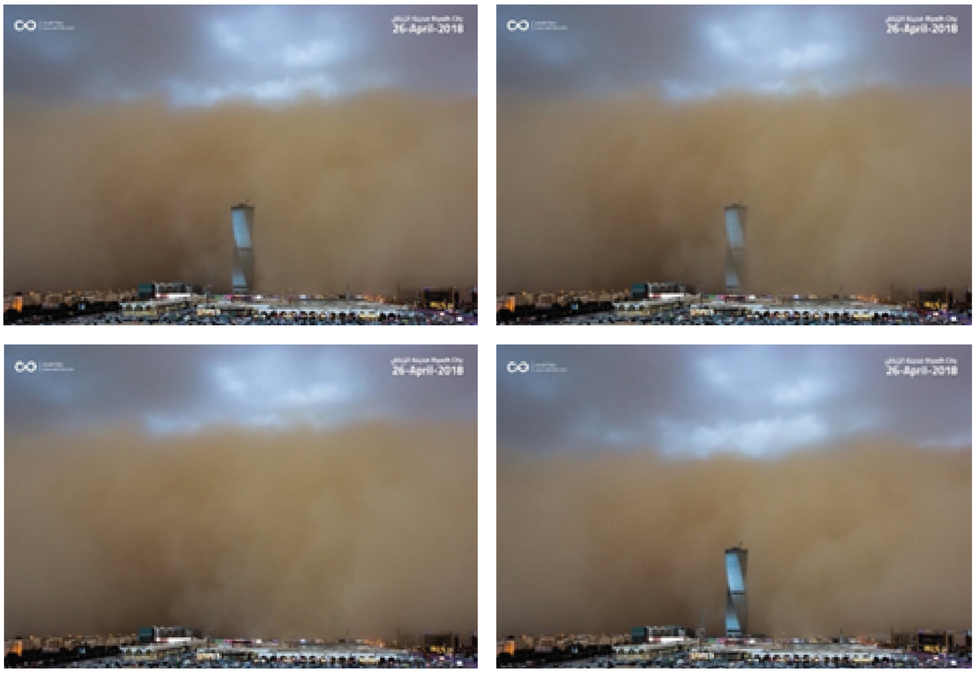


Figure 60. Heavy dust storm hits Riyadh on 26 April 2018 *(K. Mohammed [Photographer], Riyadh, 2018).*

Figure 61. Outdoor activates affected by dust storms.[1]

[1] The images are collected from my relative, through their outdoor activities.

The images presented above show that the ceremony was planned for the university stadium and was greatly influenced by the dust event. The ceremony was attended by a number of prominent dignitaries including the royal family but had to be cancelled half way through, the event was stream in different local channel. This is referred to in the RBS stage and in the entire approach where it uses actions to issue a warning regarding the expected effects of the dust event. The figures also show the effect that the storm had on a group of people enjoying a picnic. They had not been issued with an early warning of the storm that engulfed them and then faced a dangerous driving.

Evaluating the Actions in Real Events

This section reflects on the rules and actions for anticipated dust events in order to establish the degree to which these measures correspond to developments in the real world. In practical terms, the recorded weather value is entered into the prototype in order to provide an appropriate action that should be taken in the current circumstances. The choice of action and the related rules are governed by the approach set out in the methodology section.

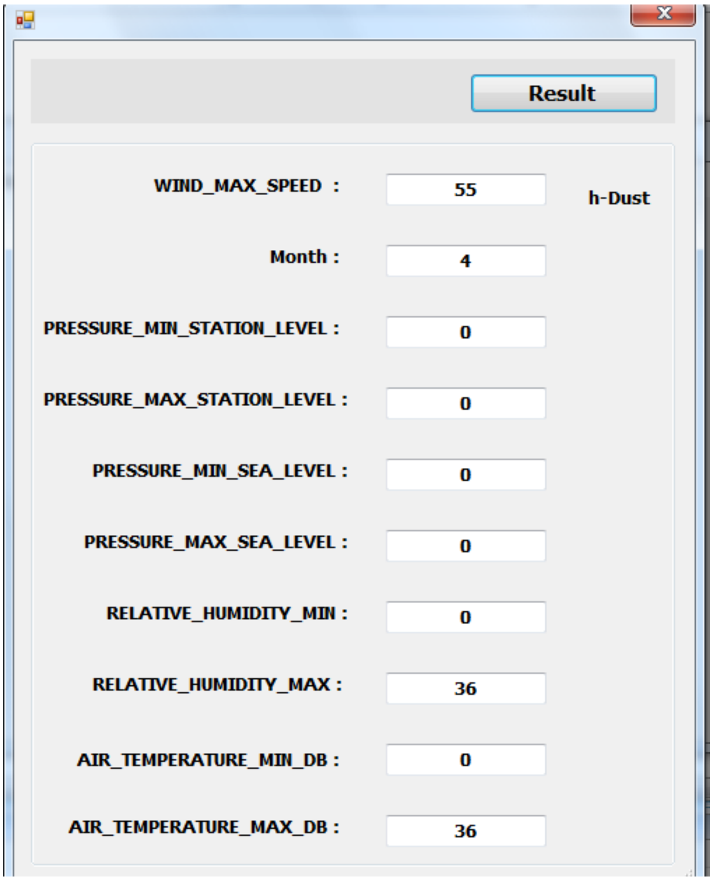


Figure 62. RBS preform for upcoming dust event.

The figure above shows the weather value being entered into the C# prototype, rules being extracted and then the CHAID algorithm being used. In this example, the rules have been applied for a high dust event which indicates that the BN-CBR prediction matched the rules.

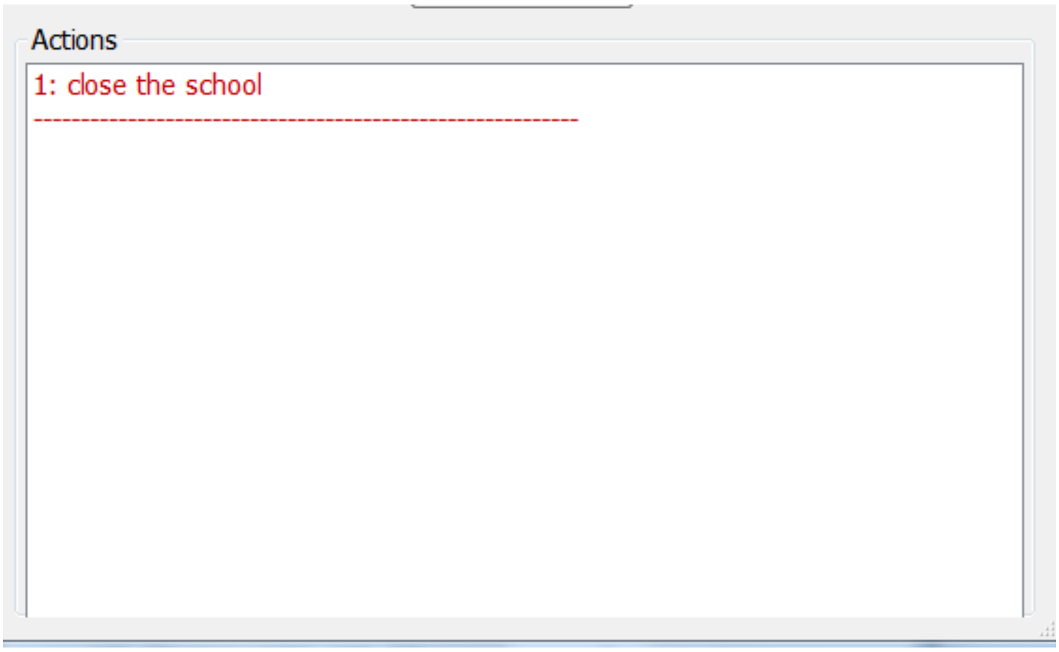


Figure 63. Action recommended for the predicted dust storm.

Based on the course of action that weather experts would recommend when faced with this particular weather scenario as well as the actions that have been taken during similar dust events of this nature, the recommendation that is issued is to close the school in response to a high dust event. When the weather experts were made aware of this recommendation, they reported as follows:

Table 21. Experts' comments on the generated safety action:

Action: High dust	Expert 1	Expert 2	Expert 3
Close the school	The chosen action is subject to the output of the various models and meteorological observations	The meteorological department and the education department must work together when deciding to close schools but based on the weather data which indicates high wind speeds, there is potential for harmful effects	If it is deemed necessary to close the school, the decision should be coordinated with the education department to ensure that the action is appropriate and does not conflict with the interests of business

The figure below shows how a dust event affecting the city of Hail on 11 May 2018 had a detrimental impact on the atmosphere, agricultural production and human health. Almisind (a weather expert) observed that the dust storm achieved P2.5 pollution per 500 micrograms which is a level that is considered likely to pose a danger to human health for those suffering with heart disease or lung problems. The adverse effects on human health were exacerbated by warnings only being issued immediately before the storm struck. It is the frequency of the condition that is fundamental to the methodology that has been chosen this work. .

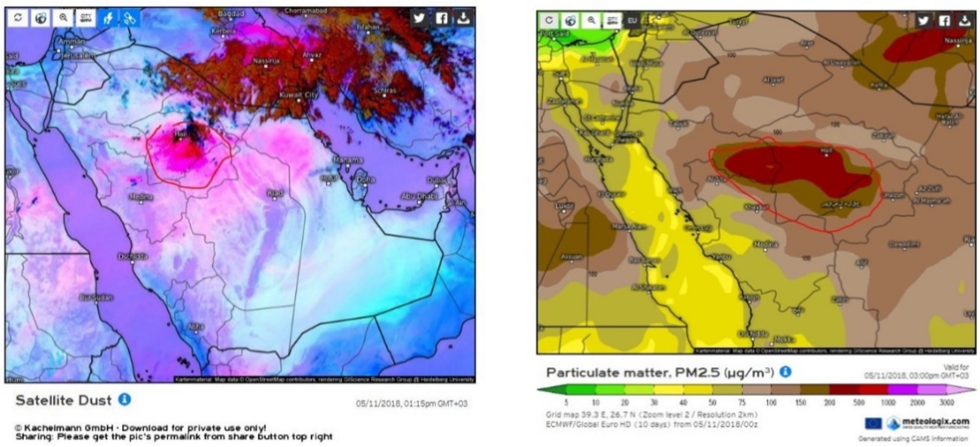


Figure 64. Hail heavy dust event on 11 May 2018 . [digital image]. Retrieved from http://meteologx.com/photos.

Hiding Weather Status in CBR

The reason for hiding 10% of the dataset was to ensure and prove with the results of the tool that BN-CBR could be properly applied in this study and perform better than pure CBR. If the result of the predication of the neighbour was accurate enough for the hide cases, that would mean the BN-CBR fit well in this study compared with pure CBR.

To do this, 10% of the final result in the dataset must be hidden. Can their neighbours predict the outcome of the cases? How efficient is BN-CBR after we do these 10 times and then average the results? The nearest neighbour should be 3 because the sample of the dust cases is quite small. As in table 14, 10 random cases were chosen from the dataset and their final results were hidden to see if their neighbours could predict the final outcomes. The results showed that CBR could predict seven correct results out of ten with high accuracy which means that BN-CBR would work perfectly for this study. The result is still the same when reducing the test to 9 iterations. In pure CBR the results showed that CBR could predict seven correct results (70%) out of ten with less accuracy than BN-CBR. The reason for this is because some of the dust cases were poorly classified, which affected the final results.

I can dispute the results that BN-CBR could be applied to predict future dust events. The results showed that there is similarity between the cases and there are many components that play more important roles than others in determine whether or not the cases involve dust which could open the door for us to search for the optimal weight of them. Finally, the prediction results were good enough to conclude that the BN-CBR approach did fit well in this thesis.

Examination and Evaluation of the Data

It is crucial that the results be examined and evaluated. This is in order to determine the effectiveness of the study after the data for old dust cases has been compiled. Generally, the system-oriented evaluation involves an internal or external evaluator which is familiar with the operation of the system and controls the evaluation through various means (Craw & Preece, 2002). First, the data are examined to ensure that they are in line with the aim of this study, which is to predict new dust cases based on old dust events. Then there is a need to determine the indicators that will be used to measure the extent to which the prediction is done. In the process, there will be formulation of a design which will incorporate these indicators. This is solely based upon the relation between the cases. In the human service delivery model of CBR, the evaluation will be addressed by giving a checklist framework to answer the queries from the new cases. The system shall take the role of responding to the queries and integrating the data in order to ensure that the most useful information is used. While testing and evaluating the proposed model, the dataset will be separated into different train tests to be determined at the time of testing and evaluation. Given that there ultimately will be a lack of rare events and homogeneity of the sets, balancing nodes of a given percentage will also be introduced. Moreover, because the dataset used in this thesis is real and historical, any predictions can easily be evaluated because the final results of the selected cases have been recorded, while the future predicted cases will be evaluated by comparing the predictions with satellite records.

Chapter Six

Conclusion and Future Work

Conclusion

The extensive research study has demonstrated that dust storms occurs when surface winds are able to lift large volumes of dust into the air, thereby reducing visibility at ground level to less than 1km. Dust storms are atmospheric events that occur as a result of the corrosion and transportation of mineral residues from the surface of the earth. Dust storms are frequent events in some regions of the world and historical experience can be used to forecast future dust events. Indeed, meteorological experts have been able to identify a number of variables that are significantly associated with dust events: low pressure, low relative humidity, high temperatures, minimal cloud cover, and high wind speeds as well as upward vertical velocity. There are various methods that can be used to forecast dust storms, and these can be divided into either passive or active methods.

However, the available methods for forecasting dust storms are not entirely accurate without using satellites image and, therefore, there remains scopes to enhance dust storm predictions. Empirical studies have suggested that new techniques need to be developed to forecast dust storms and some have even recommended that AI techniques could have a role to play. In this study the application aims to combine BN with CBR. The frequency of dust events mean that historical episodes can be used to help forecast future dust events.

There have been numerous advances in efforts to forecast dust events and mitigate their effects. One such technique that has proven to be particularly effective is CBR. This approach relies on historical cases to help solve current problems that are potentially similar in nature. Consequently, it is highly probable that the answers to the current problem are similar to those of the historical episode. This approch involves a mechanical process to retrieve historical cases that are appropriate to the current case and using this stored information to provide a solution. The problem is then revised and stored for use in the future.

In This work, a dataset comprising in excess of 7,500 cases was utilised. A BN helped to solve the uncertainty associated with CBR and enhance the forecast and clarity of the dataset before it is employed in the CBR process. Various BN structures were tested in this work in an attempt to identify the most effective model. These included the TAN, Markov and naïve Bayes structure. Each model works in a different way and this could impact on the accuracy and clarity of the results achieved. On balance, it was decided that the TAN model offered the best solution with the augmented naïve bays model being most realistic owing to the fact that there was always a degree of correlation in the data, albeit sometimes only to a small degree. In addition, the tree structure accurately reflected the causes of the dust events. Having tested the results achieved with various BN structures and observed how they fit with the CBR, the TAN was found to outperform and offer a more suitable match with CBR. CBR has been shown to offer a reliable means of identifying the characteristics of dust events and forecasting future dust events.

It has also been shown that some of the variables play a more significant role in determining whether or not a dust event occurs. For instance, it has been established that wind speed is more significant than the temperature of the air. When testing the efficiency of the CBR, 10 cases were selected at random from the dataset and their final results were not immediately revealed. Using these cases, it was found that the CBR was able to accurately forecast the outcome in 8 of these 10 cases using 3NN. In contrast, when relying solely on CBR, only 6 of the forecasts proved to be accurate. Therefore, it appears that CBR was applied appropriately in the study because the rate of forecast accuracy stood at 80%. The results confirmed that CBR offered an effective method for reliably identifying and predicting dust storms before they occur. Moreover, the RBS system was incorporated into the methodology to work in conjunction with the BN-CBR approach by extracting rules from the dust dataset and proposing appropriate actions that could help to effectively mitigate the effects of future dust events. The results achieved using RBS indicate that this approach offers the ability to accurately categorise the dust cases and propose suitable actions based on the rules that are extracted. The CHAID algorithm was used to extract appropriate rules from the dataset and generalise them in such a way that enabled a prototype expert system to be generated. There were approximately twenty rules after generalisation and weather experts were invited to offer their opinions on both the rules and the actions. Actions have been chosen that satisfy safety standards and will help to mitigate the adverse effects of dust events. Moreover, the actions have been designed to offer recommendations to three sectors of the economy: healthcare, education and the transport sector. These are the sectors that are in most need of early notification of dust events in order that they can best serve the users of these services. Schools require early notification of dust events to take mitigating action and suspend teaching if necessary. Clearly there is a need for the healthcare sector to ready equipment and increase staffing levels so that they are prepared for an increase in patients; especially those requiring attention for asthma and other breathing-related illnesses. Similarly, those responsible for the road network will need to regulate the flow of traffic in anticipation of heavy congestion at a time when visibility may be compromised by the dust storm.

The results chapter reveals that the BN (TAN) algorithm provided a good means of reclassifying in-site dust storms and reduced uncertainty in a number of cases using data for both Riyadh and Dammam with an accuracy rate in excess of 70% for classifying cases. When the BN (TAN) was combined with CBR, accuracy improved to 99% compared to the alternative combinations of CBR-ANN and CBR-CHAID that were tested.

When using the BN-CBR to forecast dust storms, the accuracy stood at 90% compared to using pure CBR. This confirms the effectiveness of matching current conditions to historical cases from the database using NN and optimal results were achieved when using 3NN. When testing RBS using data, it was found that it achieved good results and agreed with the BN-CBR predictions in the short term. RBS suggested appropriate actions for different sectors of the economy so that the effects of dust events could effectively be mitigated.

This methodology remains in its early stages of development and there remains considerable scope to extend the range of AI techniques. The available data for studies of this nature are somewhat lacking due to poor archive recordkeeping. However, this should not stand in the way of efforts to enhance AI techniques for weather forecasting and the scope for future advances remains considerable.

Contributions to Knowledge

The main contribution of this thesis is the ability to predict dust storm events based on historical dust datasets without reliance on satellite images. The results of using BN-CBR combinations to forecast dust events demonstrate that they offer an effective means of prediction with high accuracy by comparing and testing against old dust cases and evaluating with real dust examples. An additional contribution is that a prototype of an expert intelligence alarm has been built based on RBS which helps to generate a set of safety actions that could assist various government sectors to mitigate the risk of forthcoming dust events such as the health, education and transportation sectors.

The contribution is enhanced by the expert intelligence system archiving and retaining dust events so that experts will be able to study past dust cases and evaluate the effectiveness of all actions that have previously been taken to improve outcomes in future dust events.

Study Limitations

The lack archives of dust events over extended periods of time prevented more accurate and advance results being realised. In this study a real daily dataset has been applied to proposal approach, and the month attribute is indicted as key feature of dust events. it would be a great chance to test this approach in real hourly dust datasets, to expand the dust characteristics, also to predict the coming dust events in specific time scale based only on historical data. Moreover, a few safety rules and actions that are officially published or shared with metrological sector, was behind some limitation in apply and evaluate the proposed actions in this study.

Future Work

Future research into this topic would benefit from access to hourly data provided by the meteorology department. Access to this data would make it possible to forecast the precise time of a future dust event rather than merely stating on which day and time scale it will occur. This would be a significant advance in terms of accuracy and provide a more useful notification for the three sectors of the economy receiving suggested actions.

A comprehensive expert system should be developed that incorporates more effective rules to give reliable early warnings. Furthermore, the evaluated rules and actions should be expanded to cover other government sectors such as aviation. Although that would require deeper and more intensive investigations by AI specialists and the meteorology department, it would be beneficial to avoid flights being delayed and cancelled as a result of dust events. Also, generalising the approaches it would be possible to apply the methodology to other weather phenomena or other conditions that affect the atmosphere such as volcanic emissions.

In addition, the travel of dust storms from one location to another also needs to be monitored so that predictions can be made for the direction of travel and the scale of impending dust storms. This would require the existing radar stations to be relocated to enable the nature of dust events at each station location to be examined.

It would also be desirable to distribute a questionnaire among the general public in order to gauge their confidence in the ability of various forecasting applications including the methodology devised this work. This questionnaire would also help to determine whether more accurate forecasts are likely to enhance the lifestyle of Saudi citizens. In addition, it is necessary to generalise the mechanism of this methodology so that it can be applied to other weather phenomena.

References

Abuduwaili, J., DongWei, L. I. U., & GuangYang, W. U. (2010). Saline dust storms and their ecological impacts in arid regions. *Journal of arid land*, *2*(2), 144-150.

Alam, K., Trautmann, T., Blaschke, T., & Subhan, F. (2014). Changes in aerosol optical properties due to dust storms in the Middle East and Southwest Asia. Remote sensing of environment, 143, 216-227.

Al Murayziq, T. S., Kapetanakis, S., & Petridis, M. (2016). Using case-based reasoning and artificial neural networks for the efficient prediction of dust storms. *J. Expert Update*, *16*(1), 39-48.

Al Murayziq, T. S., Kapetanakis, S., & Petridis, M. (2017, December). Predicting Dust Storms Using Hybrid Intelligence System. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (pp. 338-351). Springer, Cham.

Alsaiari, N. O. (2012). An expert system for weather prediction based on animal behaviour. In The International Conference on Informatics and Applications (ICIA2012) (pp. 77-84). The Society of Digital Information and Wireless Communication.

Alves, R. R. N., & Barboza, R. R. D. (2018). Animals as Ethnozooindicators of Weather and Climate. In *Ethnozoology*(pp. 383-420).

Aood, A. S., & Jadd, N. H. (2016). STUDY SOME PROPERTIES OF LOCAL AND STORM DUST PHENOMENA IN BAGHDAD CITY. *Al-Mustansiriyah Journal of Science*, *27*(5), 31-35.

Bach, K., & Althoff, K. D. (2012, September). Developing case-based reasoning applications using mycbr 3. In *International Conference on Case-Based Reasoning* (pp. 17-31). Springer, Berlin, Heidelberg.

Ben-Gal, L., Faltin, F., & Kenett, R. (2007). Retrieved 29 December 2016, from http://www.eng.tau.ac.il/~bengal/BN.pdf.

Bergmann, R. (2000). Introduction to Case-Based Reasoning, Centre for Learning Systems and Applications, Retrieved on 22nd July 2014 from: http://www.dfki.uni-kl.de/~aabecker/Mosbach/Bergmann-CBR-Survey.pdf.

Berkat, A. (2011). Using Case-Based Reasoning (CBR) for Detecting Computer Virus, International Journal of Computer Science Issues 8(4); 606-610.

Bichindaritz, I. & Monyana, S. (2010). Case-Base Reasoning: 18th International Conference, ICCBR 2010, Alessandria, Italy, July 19-22, 2010 Proceedings. New York, NY: Springer Science & Business Media.

Biocare Medical. (2009). intelliPATH Automated Staining Instrument. [Online]. Available at: http://www.slidestainer.com/.

Black.N.T. & Ertel, W. (2011).Introduction to Artificial Intelligence. London, UK: Springer Science Publishers.

Bennacer, L., Amirat, Y., Chibani, A., Mellouk, A., & Ciavaglia, L. (2015). Self-diagnosis technique for virtual private networks combining Bayesian networks and case-based reasoning. *IEEE Transactions on Automation Science and Engineering*, *12*(1), 354-366.

BORRAJO, M. L., Baruque, B., Corchado, E., Bajo, J., & Corchado, J. M. (2011). Hybrid neural intelligent system to predict business failure in small-to-medium-size enterprises. International Journal of Neural Systems, 21(04), 277-296.

Bouckaert, R. R. (1994). Properties of Bayesian belief network learning algorithms. In Uncertainty Proceedings 1994 (pp. 102-109).

Bruland, T., Aamodt, A., & Langseth, H. (2010). Architectures integrating case-based reasoning and bayesian networks for clinical decision support. In International Conference on Intelligent Information Processing (pp. 82-91). Springer Berlin Heidelberg.

Burges, Chris, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. "Learning to rank using gradient descent." In Proceedings of the 22nd international conference on Machine learning, pp. 89-96. ACM, 2005.

Cahn, R. S. (2014). Introduction to Rule-Based Systems. J. Chem. Educ, 41(3), 116.

Cao, H., Amiraslani, F., Liu, J., & Zhou, N. (2015). Identification of dust storm source areas in West Asia using multiple environmental datasets. *Science of the Total Environment*, *502*, 224-235.

Ceccaroni, L. (2000). Integration of a rule-based expert system, a case-based reasoner and an ontological knowledge-base in the wastewater domain. In Proceedings of ECAI (p. W07).

Chen, B., Zhang, P., Zhang, B., Jia, R., Zhang, Z., Wang, T., & Zhou, T. (2014). An overview of passive and active dust detection methods using satellite measurements. Journal Of Meteorological Research, 28(6), 1029-1040. http://dx.doi.org/10.1007/s13351-014-4032-4.

Chen, S.H., Jakeman, A.J. & Norton, J.P. (2008). Artificial Intelligence Techniques: An Introduction to their Use for Modelling Environmental Systems, Mathematics and Computers in Simulation 78(2008) 379-400.

Cheng, J., & Greiner, R. (1999, July). Comparing Bayesian network classifiers. In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence (pp. 101-108). Morgan Kaufmann Publishers Inc.

Cofino, A. S., Cano, R., Sordo, C., & Gutierrez, J. M. (2002). Bayesian networks for probabilistic weather prediction. In 15th Eureopean Conference on Artificial Intelligence (ECAI).

Craw, S. & Preece, A. (2002). Advances in Case-Based Reasoning: 6th European Conference,

ECCBR 2002 Aberdeen, Scotland, UK, September 4-7, 2002 Proceedings, Volume 6. New York, NY: Springer Science & Business Media.

Dagsson-Waldhauserova, P., Arnalds, O., & Olafsson, H. (2013). Long-term frequency and characteristics of dust storm events in Northeast Iceland (1949–2011). *Atmospheric environment*, *77*, 117-127.

De Gemmis, M., Iaquinta, L., Lops, P., Musto, C., Narducci, F., & Semeraro, G. (2009). Preference learning in recommender systems. Preference Learning, 41.

de Kock, M., Le, H., Tadross, M., & Potgeiter, A. (2008). Weather Forecasting Using Dynamic Bayesian Networks. Tech. Rep., University of Cape Town.

Ding, J. (2010). Probabilistic inferences in Bayesian networks. *arXiv preprint arXiv:1011.0935*.

Dobrev, D. (Jan 19 2004). A Definition of Artificial Intelligence, Retrieved 23 July 2014 from: http://arxiv.org/pdf/1210.1568.pdf

Fielding, A. H. (1999). An introduction to machine learning methods. In *Machine learning methods for ecological applications* (pp. 1-35). Springer, Boston, MA.

Fraley, C., Raftery, A. E., Gneiting, T., Sloughter, J. M., & Berrocal, V. J. (2011). Probabilistic weather forecasting in R. The R Journal, 3(1), 55-63.

Gershenson, C. (n.d). Artificial Neural Networks for Beginners, Retrieved on 20 August 2014 from: http://arxiv.org/ftp/cs/papers/0308/0308031.pdf

Gerivani, H., Lashkaripour, G. R., Ghafoori, M., & Jalali, N. (2011). The source of dust storm in Iran: a case study based on geological information and rainfall data. Carpathian Journal of Earth and Environmental Sciences, 6.

Gillette, D. A., Adams, J., Endo, A., Smith, D., & Kihl, R. (1980). Threshold velocities for input of soil particles into the air by desert soils. *Journal of Geophysical Research: Oceans*, *85*(C10), 5621-5630.

Goudie, A. S., & Middleton, N. J. (2006). *Desert dust in the global system*. Springer Science & Business Media.

Gutiérrez Llorente, J. M., & Cano, R. (2001). Dynamic bayesian networks for rainfall forecasting. In VII Congreso Argentino de Ciencias de la Computación.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, *11*(1), 10-18.

Hamidi, M., Kavianpour, M. R., & Shao, Y. (2013). Synoptic analysis of dust storms in the Middle East. Asia-Pacific Journal of Atmospheric Sciences, 49(3), 279-286.

Haque, B. U., Belecheanu, R. A., Barson, R. J., & Pawar, K. S. (2000). Towards the application of case based reasoning to decision-making in concurrent product development (concurrent engineering). Knowledge-Based Systems, 13(2-3), 101-112.

Houeland, T. G., Bruland, T., Aamodt, A., & Langseth, H. (2010). A hybrid metareasoning architecture combining case-based reasoning and Bayesian networks.

Howes, A. L., Maron, M., & Mcalpine, C. A. (2010). Bayesian networks and adaptive management of wildlife habitat. *Conservation Biology*, *24*(4), 974-983.

Hüllermeier. E. (2007). Approximate reasoning using CBR. Berlin: Springer-Verlag. CBR versus RBR, from http://knowledgeengineering.blogspot.com/2007/12/case-based-reasoning-cbr-vs-rule-based.html Retrieved 2007.

Jamalizadeh, M. R., Moghaddamnia, A., Piri, J., Arbabi, V., Homayounifar, M., & Shahryari, A. (2008). Dust Storm Prediction Using ANNs Technique (A Case Study: Zabol City) 1.

Janssens, D. et al. (2004) Improving the performance of a multi-agent rule-based model for activity pattern decisions using Bayesian networks. Transportation Research Record: Journal of the Transportation Research Board 1984:75–83.

Jish Prakash, P., Stenchikov, G. L., Kalenderski, S., Osipov, S., & Bangalath, H. K. (2015). The impact of dust storms on the Arabian Peninsula and the Red Sea. *Atmospheric Chemistry and Physics*.

Kapetanakis, Stylianos (2012). Intelligent monitoring of business processes using case-based reasoning. PhD thesis, University of Greenwich.

Kapetanakis, Stelios & Filippoupolitis, Avgoustinos & Loukas, George & Saad Al Murayziq, Tariq. (2014). Profiling cyber attackers using Case-based Reasoning.

Kapetanakis, S., & Petridis, M. (2014). Evaluating a Case-based Reasoning Architecture for the intelligent monitoring of business workflows. In *Successful Case-based Reasoning Applications-2* (pp. 43-54). Springer, Berlin, Heidelberg.

Kapetanakis, S., Petridis, Ma, J., Bacon, L. (2010). Providing Explanations for the Intelligent Monitoring of Business Workflows Using Case-Based Reasoning. In Roth-Berghofer, T., Tintarev, N., Leake, D. B., Bahls, D. (eds*.): Proceedings of the Fifth International workshop on Explanation-aware Computing ExaCt* (ECAI 2010). Lisbon, Portugal.

Khalid, R., & Jassim, R. (2014, November). Expert diagnosis systems for network connection problems. In *Artificial Intelligence, Modelling and Simulation (AIMS), 2014 2nd International Conference on* (pp. 15-18). IEEE.

Kiskac, B., & Yardimci, H. (2004). Weather prediction expert system approaches. Ceng-568 Survey.

Kofod-Petersen, A., Langseth, H., & Aamodt, A. (2010). Explanations in Bayesian Networks using Provenance through Case-based Reasoning. In Workshop Proceedings (p. 79).

Kolodner, J. L. (1992). An introduction to case-based reasoning. Artificial Intelligence Review, 6(1), 3-34.

Kriger, C. (2007). Prediction of the influent wastewater variables using neural network theory.

Liu, H., Gegov, A., & Stahl, F. (2014). Categorization and construction of rule based systems. In International Conference on Engineering Applications of Neural Networks (pp. 183-194). Springer International Publishing.

Liu, Y., Feng, X., Wang, Q., Zhang, H., & Wang, X. (2014). Prediction of urban road congestion using a Bayesian network approach. *Procedia-Social and Behavioral Sciences*, *138*, 671-678.

Lacave, C., D´ıez, F. (2003). A review of explanation methods for Bayesian networks. The Knowledge Engineering Review 17(02) 107–127

Lopex-Fernandez, H. Fdez-Riverola, F. Reboiro-Jato, M. Glez-Pena, D & Mendez, J.R. (2011).Using CBR as Design Methodology for Developing Adaptable Decision Support Systems, Practice and Challenges From Current to Future, Retrieved on 22 July 2014 from: http://cdn.intechopen.com/pdfs-wm/19337.pdf.

Main, J., Dilon, T. &Shiu, S. (2001). A Tutorial on Case-Based Reasoning, Soft Computing in Case Based Reasoning, pp. 1-28.

Mascarenhas, S. (2010). Case Based Reasoning & Fame IA, Retrieved on 22 July 2014 from: https://fenix.tecnico.ulisboa.pt/downloadFile/3779574683590/ASS-Case-Based%20Reasoning.pdf.

Massie, S., Craw, S., & Wiratunga, N. (2007, August). When similar problems don't have similar solutions. In *International Conference on Case-Based Reasoning* (pp. 92-106). Springer, Berlin, Heidelberg.

McDermott, J. (1982). R1: A rule-based configurer of computer systems. Artificial intelligence, 19(1), 39-88.

Merli, M., Moscatelli, M., Mariotti, G., Pagliaro, U., Bernardelli, F., & Nieri, M. (2016). A minimally invasive technique for lateral maxillary sinus floor elevation: a Bayesian network study. *Clinical oral implants research*, *27*(3), 273-281.

Modeler, I. S. (2012). Algorithms Guide. IBM Corporation.

Moradian, M. J., Rastegarfar, B., reza Rastegar, M., & Ardalan, A. (2015). Tehran dust storm early warning system: corrective measures. *PLoS currents*, *7*.

Möller, A., Lenkoski, A., & Thorarinsdottir, T. L. (2013). Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, *139*(673), 982-991.

Najafi, M. S., Khoshakhllagh, F., Zamanzadeh, S. M., Shirazi, M. H., Samadi, M., & Hajikhani, S. (2014). Characteristics of TSP loads during the Middle East springtime dust storm (MESDS) in Western Iran. Arabian Journal of Geosciences, 7(12), 5367-5381.

Nikpour, H., Aamodt, A., & Skalle, P. (2017). Diagnosing root causes and generating graphical explanations by integrating temporal causal reasoning and CBR. CEUR Workshop Proceedings.

Onishi, K., Otani, S., Yoshida, A., Mu, H., & Kurozawa, Y. (2015). Adverse health effects of Asian dust particles and heavy metals in Japan. *Asia Pacific Journal of Public Health*, *27*(2), NP1719-NP1726.

Otim, S. (2006, May). A case-based knowledge management system for disaster management: fundamental concepts. In *Proceedings of the 3rd International ISCRAM Conference* (Vol. 2006, p. 598604). Newark, NJ (USA).

Padmanaban, H. (2014). Comparative analysis of Naive Bayes and tree augmented naïve Bayes models.

Papayannis, A., Amiridis, V., Mona, L., Tsaknakis, G., Balis, D., Bösenberg, J., ... & Mitev, V. (2008). Systematic lidar observations of Saharan dust over Europe in the frame of EARLINET (2000–2002). *Journal of Geophysical Research: Atmospheres*, *113*(D10).

Panda, M., Abraham, A., & Patra, M. R. (2012). A hybrid intelligent approach for network intrusion detection. Procedia Engineering, 30, 1-9.

Pearl, J. (2000) Causality: Models, Reasoning, and Inference. Cambridge University Press.

Prentzas, J., & Hatzilygeroudis, I. (2009). Combinations of case-based reasoning with other intelligent methods. *International Journal of Hybrid Intelligent Systems*, *6*(4), 189-209.

Priddy, K.L & Keller, P.E. (2005). Artificial Neural Networks: An Introduction. Bellingham, DC: SPIE Press.

Pavón, R., Díaz, F., Laza, R., & Luzón, V. (2009). Automatic parameter tuning with a Bayesian case-based reasoning system. A case of study. *Expert Systems with Applications*, *36*(2), 3407-3420.

Pye, K. (2015). *Aeolian dust and dust deposits*. Elsevier.

Qian, Z. A., Shong, M. H., & LI, W. Y. (2013). Forecasting and early warning of dust storms. Journal of Desert Research, 2, (pg 1-14).

Richter, M. M., & Weber, R. O. (2016). *Case-based reasoning*. Springer-Verlag Berlin An.

Root, H. E. (2012). Probability statements in weather forecasting. Journal of Applied Meteorology, 1(2), 163-168.

Roth-Berghofer, T., Goker, M.H., Guvernir, H.A. (2006).Advances in Case-Based Reasoning: 8th European Conference, ECCBR 2006, Fethiye, Turkey, September 4-7, 2006 Proceedings. Berlin, Germany: Springer-Verlag Publishers.

Roth-Berghofer, T. & Bahls, D. (2008). Explanation Capabilities of the Open Source Case-Based Reasoning Tool myCBR, in Petridis, m. & Wiratunga, N (eds). (2008). Proceedings of the Thirteenth UK Workshop on Case-Based Reasoning. UK, London: University of Greenwich.

Schiaffino, S. N., & Amandi, A. (2000). User profiling with Case-Based Reasoning and Bayesian Networks. In IBERAMIA-SBIA 2000 open discussion track (pp. 12-21).

Sharma, V., Rai, S., & Dev, A. (2012). A comprehensive study of artificial neural networks. *International Journal of Advanced research in computer science and software engineering*, *2*(10).

Shiu, S.C.K. & Pal, S.K. (2004). Case-Based Reasoning: Concepts, Features and Soft Computing, Applied Intelligence, 21, 233-238.

Sissakian, V., Al-Ansari, N., & Knutsson, S. (2013). Sand and dust storm events in Iraq. Natural Science, 05(10), 1084-1094. http://dx.doi.org/10.4236/ns.2013.510133

Stephenson, T. A. (2000). An introduction to Bayesian network theory and usage (No. EPFL-REPORT-82584). IDIAP.

Терентьєв, О. М., Кириченко, В. Е., Связінська, Н. О., & Просянкіна-Жарова, Т. І. (2016). Financial Risk Forecasting Using Naïve and Tree Augmented Classifier Based on Bayesian Networks. *Naukovi Visti NTUU KPI*, (2), 60-68.

Tong, D. Q., Wang, J. X., Gill, T. E., Lei, H., & Wang, B. (2017). Intensified dust storm activity and Valley fever infection in the southwestern United States. *Geophysical Research Letters*, *44*(9), 4304-4312.

Tsamardinos, I., Brown, L. E., & Aliferis, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. Machine learning, 65(1), 31-78.

Tran, H., Schönwälder, J. (2008). Fault Resolution in Case-Based Reasoning. In: Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence, Springer 429.

UNEP, WMO & UNCCD (2016). Global Assessment of Sand and Dust Storms. United Nations Environment Programme, Nairobi, 2(4), 281-288. http://dx.doi.org/10.1016/0002-1571(65)90014-2.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. Ecological modelling, 203(3-4), 312-318.

Veenendaal, A., Daly, E., Jones, E., Gang, Z., Vartak, S., & Patwardhan, R. S. (2015). Decision Rule Driven Human Activity Recognition. *Computer Science and Emerging Research Journal*, *3*.

Vidal, J. (2009). Dust storms spread deadly diseases worldwide. *the guardian*.

Wang, N., Yao, T., Yang, X., Shen, J., & Wang, Y. (2007). Variations in dust event frequency over the past century reflected by ice-core and lacustrine records in north China. Science In China Series D: Earth Sciences, 50(5), 736-744. http://dx.doi.org/10.1007/s11430-007-0037-5.

Wei, H., & Tsatsoulis, C., (1993). Uncertain Case-Based Reasoning. In Automated Reasoning (pp. 169). Springer Netherlands.

Wilamowski, B. (2002). Neural networks and fuzzy systems. *The Microelectronic Handbook*.

Wong, S. K. M., & Xiang, Y. (1994). Construction of a Markov network from data for probabilistic inference. In *Third International Workshop on Rough Sets and Soft Computing* (pp. 562-569).

Xiong, W., Zhang, Y., & Yin, C. (2009). Optimal energy management for a series–parallel hybrid electric bus. *Energy conversion and management*, *50*(7), 1730-1738.

Zhang, H. (2004). The optimality of naive Bayes. AA, 1(2), 3.