International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

An Evaluation of Machine Learning and Big Data Analytics Performance in Cloud Computing and Computer Vision

Hather Ibraheem Abed[1], Nawar A. Sultan[2], Osama Yassin Mohammed[3]
[1]Northern Technical University, Iraq
Email: hadiria@ntu.edu.iq
[2]Northern Technical University, Iraq
Email: nawarabd@ntu.edu.iq
[3]Northern Technical University, Iraq
Email: osama.yassin@ntu.edu.iq

Abstract: Although cloud computing is receiving a lot of attention, security remains a significant barrier to its general adoption. Cloud service users frequently worry about data loss, security risks, and availability issues. Because of the accessibility and openness of the huge volume of data amassed by sensors and the web throughout recent years, computer applications have seen a remarkable change from straightforward data processing to machine learning. Two widely used technologies, Big Data and Cloud computing, are the focus of worry in the IT industry. Enormous data sets are put away, handled, and broke down under the possibility of "Big Data." Then again, cloud computing centres around giving the framework to make such systems conceivable in a period and cash saving way. The objective of the review is to survey the Big Data Analytics and Machine learning ideal models for use in cloud computing and computer vision. The programmed data examination of enormous data sets and the production of models for the wide connections between data are the centre highlights of machine learning (ML). The usefulness of machine learning-based strategies for identifying threats in a cloud computing environment is surveyed and compared in this research.

Keywords: Machine learning, Artificial intelligence, Big data analytics, Cloud computing, Computer vision.

I. INTRODUCTION

The use of cloud computing has grown in popularity recently. The customised data centres have gained popularity as a low-cost infrastructure option for corporate strategies. Internet services are just one of the many resources that cloud computing provides. Cloud computing offers a variety of online resources that help individuals and organisations cut infrastructure costs. End users continue to utilise and distribute infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS) in large numbers. In this approach, customers can deploy their apps without needing to maintain or control the infrastructure for cloud computing or have any understanding of it. Instead, they merely use hardware or software that they access or rent and only pay for what they really use. The option to pay as you go, which is highly sought by cloud hosting providers, is becoming more and more common in the corporate computing paradigm (KOBIELUS, 2018).

The driving factor behind all of the earth's sciences will be big data. Massive amounts of data are being produced by the digital empire (in Zetta Bytes now). All areas of science, particularly astronomy and medical research, primarily rely on data collected from a variety of sources. In the current technological environment, producing data is not a problem, but processing, storing, and retrieving it is a major challenge. The issue of learning, analytics, and prediction from such a vast volume of data is even more serious. But thanks to cloud computing, the scenario for big data analysis and storage is no longer as difficult to complete as it once was.

The objective of machine learning and computer vision is to enable computers to detect data, appreciate data, and make moves in light of past and present results. Computer vision and machine learning research are still being developed. The Web of Things, Modern Web of Things, and cerebrum human points of interaction all rely upon computer vision. Utilizing machine learning and computer vision, the complicated human ways of behaving in media streams are distinguished and monitored. There are various dependable procedures for examination and expectation, including semi-regulated, solo, and managed learning. These methods utilize machine learning calculations like help vector machines and KNN, among others (Dash, 2013).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

1.1. Background of the study

A few significant subjects, including computer vision and machine learning, have been the focal point of ongoing review. The computer vision framework utilizes picture and example mappings to find arrangements. An assortment of pixels is the way it sees a picture. Using computer vision, checking, investigation, and observation errands are robotized. Machine learning is a subset of artificial intelligence. The programmed explanation and examination of recordings is an outcome of computer vision and machine learning.

The three methods used in machine learning and computer vision are directed, unaided, and semi-regulated learning. Administered learning was utilized to mark the preparation data. The course of data naming is pricy, tedious, and work concentrated. Be that as it may, with semi-directed learning, a portion of the data is named and some isn't. The Bayesian organization classifier offers the benefit for learning from unlabeled data. Most of issues in the genuine world, be that as it may, fall inside the classification of solo learning, where examples arise because of grouping. The computer vision machine learning standards incorporate help vector machines, brain organizations, and probabilistic graphical models. Support vector machines (SVMs), a part of directed machine learning procedures, are a well known order technique (HASHEM, 2015).

Enormous measures of data are produced in the Web of Things (IoT) age and are obtained from different heterogeneous sources, including cell phones, sensors, and virtual entertainment. Big Data faces huge difficulties with regards to handling, stockpiling, and scientific capacities. Cloud computing gives a valuable and savvy elective for supporting the stockpiling of Big Data and the activity of data insightful applications. In the Internet of Things, artificial intelligence (AI) is used for data analytics and mining, and a cloud computing environment is used for data processing and data exchange.

1.2. Problem Statement

The parallel distributed computing system known as "cloud computing" is becoming a popular computing tool for big data analytics. However, neither approach deals with the complexity in space and time constraints. Since computer worms and viruses, which are intelligent agents, are responsible for the majority of network-centric cyberattacks, it is fundamental for battle them with astute semi-independent specialists that can perceive, survey, and respond to digital assaults. The majority of organisations now must update their cyber defence strategy due to the quick development of computing and digital technology. Subsequently, security network directors should be lither, adaptable, and give solid digital protection frameworks to the speedy ID of online dangers. Assessing machine learning (ML) and big data analytics (BDA) ideal models for use in digital protection is the main pressing concern.

1.3. Objectives of the study

The following are the goals of this research study:

- To research and critically assess big data and machine learning applications in computer vision.
- To analyze the applications of big data analytics and machine learning in cyber security.

1.4. Research Questions

The primary questions for research are:

Q1: What new solutions (hardware or software) are being developed for effective data analytics and machine learning in cloud systems of the future?

Q2: Which Big Data and Machine Learning paradigms are best for creating a cyber security system?

1.5. Significance of the study

According to present circumstances, a data breach could happen if extreme precautions are not taken beforehand. Big data and machine learning in general, as well as the use of cloud-based services, models, architectures, technologies, and cyber security to support their effective use in terms of processing, analytics, prediction, inference, and intelligence We talk about different cloud-based specialist services that are focused on different facets of data science and machine learning.

II. Review of Literature

When patients, healthcare professionals, and other parties communicate sensitive information, Abdul Majeed discusses the necessity for data privacy. The proposed solution attempted to provide data anonymization even when the hackers were aware of the backdrop, thereby differing from the current methods. Data is processed into intervals with preset lengths, and the original value is replaced with average values. The suggested method offers confidentiality and privacy of healthcare data throughout data publishing together with successful simulation outcomes (Majeed, "Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data", 2019).

In-depth explanations of data security and privacy are provided by Razak et al.. When a data breach occurs and the privacy of the data is called into question, the identifier that is used exclusively for the data holder may be at danger.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

Additionally, when data is transferred from one party to another, it is penetrated, exposing the data and opening it up to modification by attackers, such as via forgeries or spoofing. The suggested approach uses the vein in the palm for biometric authentication, followed by database anonymization to adequately protect the data. Additionally, the system makes sure that user data is not exposed when data is leaked (S. Abd Razak, 2020).

What big data is, how people and organisations might use it, what problems might follow if consumption keeps growing, and other topics are clearly discussed by Sagiroglu et al. Also covered in detail are the privacy and security issues. The difficulties that may occur when big data is used by businesses are described in detail. Big data platforms are also mentioned, which are highly beneficial for the industries in the present (S. Sagiroglu and D. Sinanc, 2013).

To identify intruders, Tsai C et al .'s approach based on machine learning is shown. Utilizing the internet exposes you to the possibility of network assaults. Unauthorized entry into networks is a primary form of network attack that needs to be stopped in order to protect the network. Intruders typically act differently than regular users. the methods of machine learning that can be used to secure a network and assess the behaviour of illegal users (Tsai C, 2009).

Using machine learning, Karn RR et al. explore the automatic tweaking of models. A distributed and dynamic computing framework is the cloud. The models based on machine learning techniques may not be adequate for monitoring cloud data due to its variable nature. When characteristics of the input data change, the accuracy of the model may alter over time. It also demands extensive training and resources. This study suggests a cloud DevOps architecture with automatic tuning and selection that generates dynamic instantiations (Karn RR, 2019).

A Collaborative Anomaly Detection Framework (CADF) was recommended by Moustafa et al. to manage huge data in cloud computing. Give technical assistance and instructions on how to use this framework in these locations. The recommended approach comprises of three modules: assembling and logging network data; pre-handling this data; and another choice motor that distinguishes attacks by using the Gaussian Blend Model and the lower-upper Interquartile distance limit. The novel Decision Engine was put to the test by being compared to three ADS techniques while modelling on actual cloud computing systems using the UNSW-NB15 database. This mode's Software as a Service (SaaS) design makes it simple to install in cloud computing (Moustafa N, 2017).

Cloud-based anomaly detection was first described by Mobilio et al. as a service that declares control over the idea of inaccurate discovery by means of a standard rule utilised in cloud systems. Moreover, they offer primer discoveries utilizing lightweight machines that indicate a feasible method for working on the possibility of deformity distinguishing proof. They likewise discussed how to transform the idea of troublesome obtaining into a mysterious securing administration by utilizing the as-administration model. Moreover, they prompt fostering a worldview that can help out any review framework that keeps data in a succession of time series and supports worldview as a help. Results from fundamental testing of as a help with the Clearwater cloud framework show how the as-a-administration worldview can oversee discovery rationale effectively. This strategy is intriguing since it makes use of cutting-edge technologies for novel real-time detection (Mobilio M, 2019).

Utilizing an unequivocal ML model, Zhang proposed multi-view learning procedures for distinguishing failures in cloud computing stages. They work with a real-time gap that is formed by the simultaneous occurrence of two phases and is trained using various ELM model features. By minimising training errors, the presented approach automatically combines several features from various sub-systems and finds a better separation solution. The link between the samples and the separation boundary indicates conflict calculated between Sum, and the weighted samples determine the separation model's recurrence rate. The suggested model does well with Multi-view learning and feed control, but it confronts a number of difficulties in detecting errors, such as distribution imbalances, high-magnitude features, etc (Zhang J, 2019).

The major areas of interest for Kaisler et al. include the origins of big data, its growth, and how businesses and consumers use it. Once big data is being used, a number of challenges could occur, including storage, issues with data transmission, issues with data processing, and problems with data management. Along with the techniques that are accessible for big data analysis, the authors also discussed the importance of using big data when real-time applications are in use. It also provides a thorough grasp of the big data difficulties, the reasons behind big data analysis, and design methodologies (Stephen Kaisler, 2013).

III. RESEARCH METHODOLOGY

Focus group discussions are one of the qualitative research methods used by the researcher. Since there is no need to quantify the investigation, it is done to identify discrepancies between data analytics models for cyber security. This study used a case study research methodology. In this regard, every data analytics model for cyber security is treated as a distinct

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

instance to be looked into in a context all of its own. Case studies have frequently been used in earlier studies of cyber security. To allow for comparison, the researcher creates a control case that takes into consideration the optimum data analytics approach for cyber security.

3.1. Research Design

In order to comprehensively characterise current realities and characteristics of big data analytics models for cyber security, the researcher uses a descriptive study design. The study's main goal is to describe the models in great detail.

3.2. Tools for data collection

In a research process, the choice and preparation of instruments and technologies for data collecting is crucial. The effectiveness and applicability of a research tool may have a significant impact on the study's conclusion. Using the proper statistical methods, the data gathered from the secondary sources was processed and examined. The investigation was carried out in a way that highlighted variances in several facets of tribal social development. The necessary information was gathered using the main anthropological techniques, including interviews, case studies, and observation. The data was collected using the questionnaire study method. The qualitative exploratory method was adequate given the short time frame. The sample size was constrained as a result of this limitation, making it difficult to sum the outcomes. A questionnaire review tool was used to acquire data.

3.3. Technique for Collecting Data

3.3.1. Primary data: Data gathering based on questionnaire-derived primary data.

3.3.2. Secondary data: Through past fieldwork, pertinent books, journals, census data, and reports, secondary data was acquired.

IV. DATA ANALYSIS

The researcher refers to the qualities of a perfect data analytics model for cyber security when examining the various data analytics models for cyber security. The specialist integrates many literature sources when creating an ideal model. Big data, analytics, and insights make up the three main building blocks of the fundamental big data analytics approach for cyber security. But a fourth element—prediction (or predictive analytics)—could be mentioned.

4.1. Big data analytics

More solid big data analytics models for network safety have been made in data mining and machine learning ways to deal with take care of big data's digital protection issues. Data mining reactors and calculations, malware and interruption recognition strategies, and vector machine learning procedures are completely utilized in big data analytics for network protection. Contrarily, it has been shown that hostile programmes frequently change their behaviour in order to evade the reactors and algorithms created to find them. Additionally, unbounded patterns, data non-stationarity, inconsistent delays, uniqueness, high misleading problem rates, and conspiracy assaults provide difficulties for intrusion detection systems. Big data analytics for cyber security hence require a multi-layered and multi-dimensional methodology. As such, a big data analytics model for cyber security that is effective must be able to identify malware and intrusions at every level of the security architecture.

4.2. Computer vision with machine learning

The review researched various machine learning applications in computer vision. Highlight extraction, design coordinating, structure portrayal, surface remaking, and displaying for organic sciences are a couple of instances of how to section data and refine visual models. The translation of data from pictures containing vehicle and passerby identification, the programmed characterization of rail line tie deficiencies utilizing pictures, the separation of mango assortments in light of size ascribes, and the extraction of graphical and literary data from record pictures all utilization machine learning in computer vision. Other similar applications include face and gesture recognition, machine vision, the ability to read handwritten characters and numerals, upgraded driver help frameworks, social examinations, and position assessment. Finding control slopes in Google Road View can be done by automatically spotting them and looking them up in pictures. Imaginative purposes for machine learning and computer vision can be tracked down in designing, medication, farming, cosmology, sports, and training, among different fields (V. Dhar, 2013).

4.3. Big Data Analytics in Cloud Computing

To give adaptable assets, accelerate development, and accomplish economies of scale, cloud computing is the dissemination of computing administrations like waiters, stockpiling, databases, organizing, programming, analytics, and so on through the Web (the "cloud"). How computing foundation is preoccupied and utilized has been changed by cloud computing. Everything that may be thought of as a service is now included in cloud paradigms. The different benefits of cloud computing, including its flexibility, pay-more only as costs arise or pay-per-use plan of action, low starting expense of capital, and so on, have made it an alluring and viable choice for big data the executives and analytics. Big data is already regarded as essential for many businesses

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

and industries, thus service providers like Amazon, Google, and Microsoft are now delivering their own big data frameworks at a reasonable price. All sizes of businesses can scale these systems. As a result, the phrase Analytics as a Service (AaaS) became well-known as a quicker and more effective approach to connect, change and visualise various sorts of data (M. I. Jordan, 2015).

4.4. Predictive analytics

In order to determine the possibility of a cyber security event occurring in the future using current cyber security data, a big data analytics model for cyber security is applied in predictive analytics. To be successful in social occasion big data about network safety, breaking down big data about digital protection dangers, giving significant experiences, and estimating potential future network protection occasions, a data analytics model for network protection should have the option to coordinate these components.

V. RESULT AND DISCUSSION

The capacity to classify and store them as per the novel characteristics is a work concentrated task, not at all like the text based material, which is continually moving in the web-based world. It takes computer mediations with cutting edge model-based vision abilities and master capacity to list and store graphical data. The exploration on machine learning and big data analytics in a few fields is featured in this paper. The expense, exertion, and time in designing, research, and technology have all been decreased because to machine learning approaches. The detection of human emotions is mechanised utilizing machine learning and computer vision (likes and dislikes confidence levels). The probabilistic models use naming and example acknowledgment to expect human way of behaving. In elite athletics, player and group execution is estimated and dissected utilizing machine learning and computer vision. Additionally, it has been utilised in other sectors for preventative maintenance. The viability and proficiency of the assembling units are fundamentally affected by supplanting hardware and apparatuses in enterprises before they break down. A critical wellspring of data is the public camera framework and savvy contraptions with sensors. At the point when these data are exposed to computer vision and machine learning methods, it is feasible to expect and follow city traffic.

Data mining algorithms need to be improved and optimised for classification of intrusion attacks. Table 1 below lists the advantages and disadvantages of every calculation utilizing the NSL-KDD dataset.

Table: 1. Performance of Decision Tree Algorithms, K-Nearest Neighbour, Naive-Bayes, Artificial Neural Networks, and Support Vector Machines

Parameter	SVM	ANN	KNN	NB	DT
Incidents classified properly	35621	35225	36153	33672	36183
Incidents that have been misclassified	775	2171	243	3724	213
Kappa Statistics	1.9664	1.9238	1.9986	1.8008	1.9923
Mean Absolute Error	1.0369	1.0647	1.0158	1.1136	1.0166
Root Mean Squared Error	1.1736	1.298	1.0850	1.3254	1.0753
Relative Absolute Error	6.3778%	12.119%	2.1435%	21.7919%	2.2956%

5.1. Support vector machine

Support Vector Machine is a machine learning and artificial intelligence classification technique that uses a set of points in X-dimensional space that are divided into two kinds. Utilizing direct part or nonlinear bit works, the help vector machine makes a (X — 1) layered hyper plane for gathering these focuses into at least two classifications. A technique for characterizing bank execution into four gatherings of solid, good, ordinary, and terrible execution is given by piece capabilities to polynomial, spiral, and multi-facet perceptual classifiers. The role of the function determines the class of bank performance.

$$Performance\ class = f(\vec{x}.\vec{w}) = f\left(\sum_{j} \quad x_j w_j\right)$$

Table: 2. Performance of Support Vector Machine

sigma	c	Accuracy	Kappa	Accuracy SD	Kappa SD
0.0524110	0.27	0.803425	0.698738	0.0976110	0.160514
0.0524110	0.52	0.796209	0.681556	0.089888	0.152754
0.0524110	1.00	0.811593	0.698896	0.0823511	0.146668

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

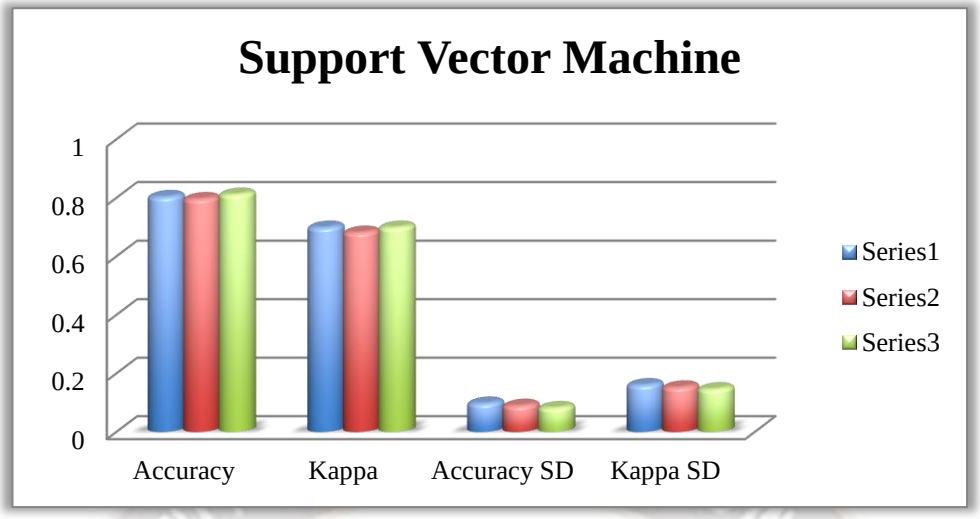

Figure: 1. Graphical representation of SVM performance

As indicated by Table 2, the SVM model's exactness rate for guaging bank dissolvability on the preparation dataset was 81.3%. Our incredibly compelling model's ideal tuning sigma and cost values were 0.07 and 1. The Kappa measurement and the Kappa SD were, separately, 69.11% and 0.15. The technique had a precision level of 94.7% and a kappa of 90.56% on the test dataset.

5.2. KNN algorithm

With the aid of the training dataset, the K-NN algorithm, a non-parametric supervised machine learning method, endeavours to order a data point into specific classifications. By looking through the entire training dataset for the K most similar instances or neighbours, predictions are made for a new item (y). In order to accomplish this, the method determines the Euclidean distance as follows:

$$d(x,y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}$$

Table: 3. Performance of KNN Algorithm

K	Accuracy	Kappa	Accuracy SD	Kappa SD
7	0.6190667	0.3899953	0.1482398	0.23601210
9	0.6470886	0.42749410	0.1766942	0.2905526
11	0.68239910	0.4917578	0.1949925	0.3083392

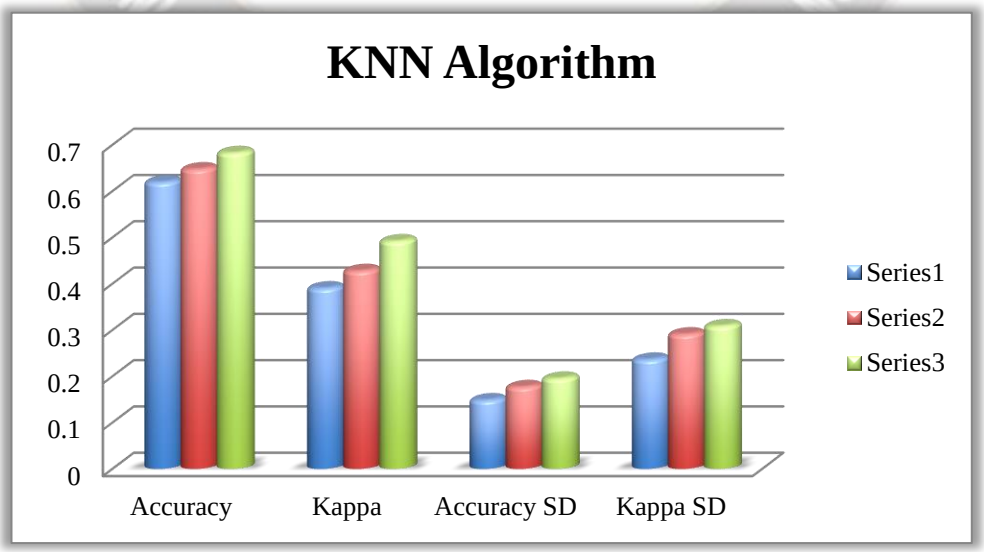

Figure: 2. Graphical representation of KNN algorithm performance

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

The training dataset's accuracy rate was 68.4%. The Kappa measurement was 49.4%, and the Kappa SD was 0.19. The calculation's precision level and kappa on the test dataset were both 69.7% and 51%, individually.

5.3. Multi Linear Discriminant Analysis (LDA)

A method for reducing the number of dimensions is linear discriminant analysis. Dimensionality reduction commonly referred to as the path of dimensionality, is a method of lowering the number of random variables taken into account by identifying a group of principal variables. The LDA determines the between-class variance, also known as the separability between n classes. D_b can represent the separation between n classes.

$$D_b = \sum_{i=1}^{g} N_i \left(\underline{x_i} - \underline{x}\right)\left(\underline{x_i} - \underline{x}\right)^T$$

Table: 4. Performance of Linear Discriminant Algorithm

Accuracy	Kappa	Accuracy SD	Kappa SD
0.82444111	0.7238333	0.1218838	0.179509

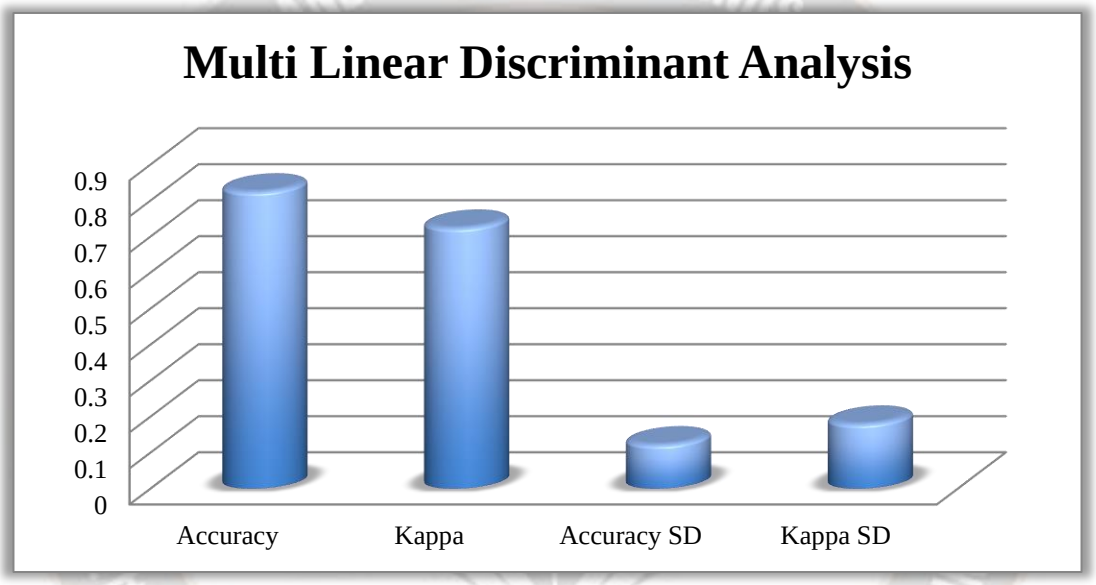

Figure: 3. Graphical representation of LDA performance

As displayed in table 4, the LDA achieved a precision level of 82% on the preparation dataset. The Kappa SD was 0.18 and the Kappa measurement was 72%. The calculation accomplished a kappa of 86.66% and a 92% precision level on the test dataset.

5.4. Classification and Regression Trees (CART)

The exhibition aftereffects of our Truck calculation in estimating bank disappointment on the preparation set are shown in Table 6 underneath.

Table: 5. Performance of the CART model

Complexity Parameter	Accuracy	Kappa	Accuracy SD	Kappa SD
0.07051337	0.8477114	0.77215110	0.051784710	0.07274594
0.17773627	0.7985172	0.68852410	0.07922918	0.16059944
0.44485955	0.5424366	0.1350613	0.08385373	0.20752624

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

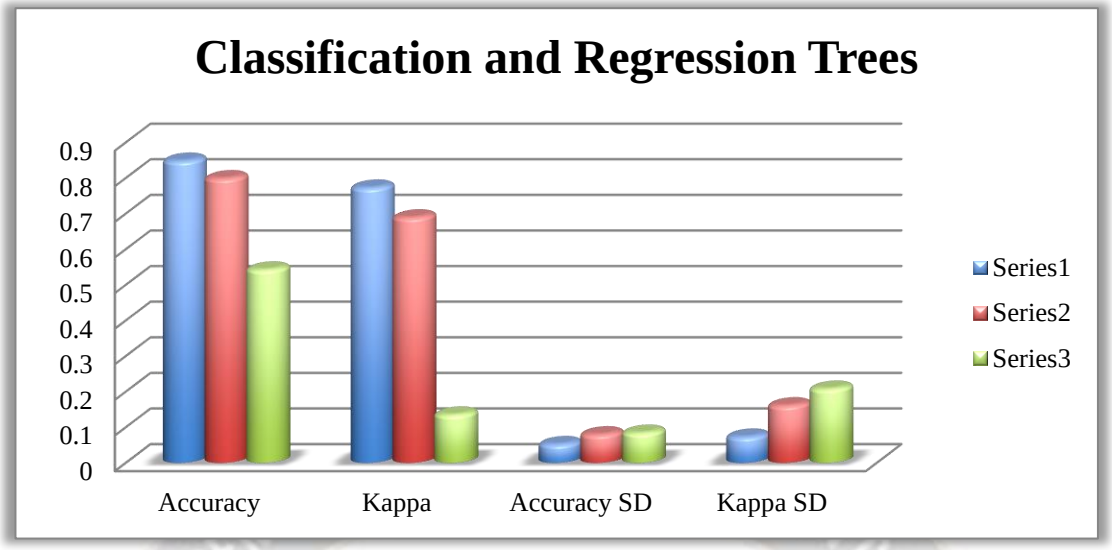

Figure: 4. Graphical representation of CART performance

On the preparation dataset, the calculation's exactness rate was 83.10%. Our ideal model's best tuning or intricacy boundary was 1.070. Our classifier was compelling, as shown by the Kappa measurement of 77% and the Kappa SD of 0.07 in the characterization of bank classifications. The technique had a kappa of 89.74% and an exactness level of 94.7% on the test dataset.

5.5. Random Forest Classifier

An outfit approach utilized for arrangement and relapse issues is the Arbitrary Timberland classifier. A haphazardly picked subset of the preparation data is utilized to create an assortment of choice trees. The ultimate class of an instance in the classification issue is then decided by adding the votes from various decision trees. The impact of trees with low error rates is increased because the trees with greater mistake rates are given less weight relative to other trees.

Table: 6. Random Forest Performance

mtry	Accuracy	Kappa	Accuracy SD	Kappa SD
4	0.8474549	0.7623442	0.12416656	0.17440281
16	0.8756234	0.8031913	0.06271738	0.09505152
18	0.8684806	0.7920957	0.066582610	0.09983993

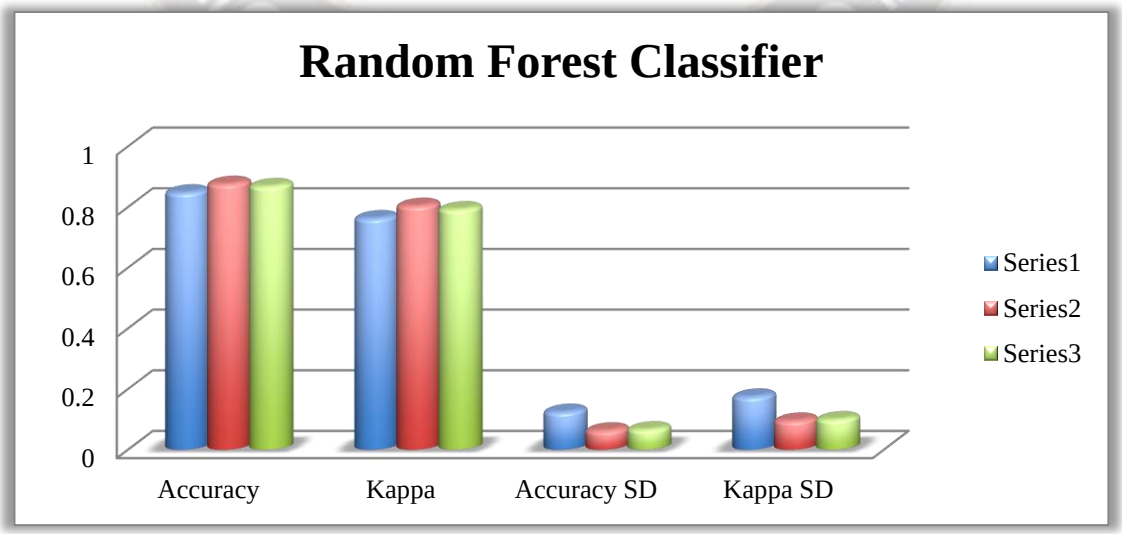

Figure: 5. Graphical representation of Random forest performance

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7144
***Article Received**: 25 March 2023 **Revised**: 30 April 2023 **Accepted**: 08 June 2023*

Our arbitrary woodland's exactness on the preparation set was 87.7%, as displayed in table 6. The quantity of indicators decided indiscriminately to construct trees was the ideal tuning boundary for our model, and it was 16. The Kappa measurement was 80.5%, and the Kappa SD was 0.11, separately. The calculation's exactness level and kappa were both 98% on the test dataset. Contrasted with different techniques, the calculation was very successful at arranging bank execution.

The research findings interprets that the application of machine learning with big data brings out the best combination and this can assess the computer visions and performs exceptionally in field of cyber security. Hence, the proposed objectives have been fulfilled by the research implementation.

VI. CONCLUSION

The improvement of novel techniques, systems, models, and calculations is a consequence of the rising business and scholarly examination on cloud computing and computer vision. Many problems with feature extraction and processing in computer vision have been solved via machine learning. Depending on the domain, the outputs of computer vision's machine learning applications vary. The application of machine learning and big data in computer vision is analysed, categorised, and discussed in this work. As a subset of artificial intelligence, machine learning algorithms can be categorised as supervised, unsupervised, semi-supervised, and reinforcement learning algorithms. Large data sets are automatically analysed for patterns and relationships, which leads to the creation of models for those patterns. Big data analytics centres around the volume, variety, and speed of data notwithstanding the size of the actual data. Big data is described by its volume, velocity, variety, veracity, veracity, vocabulary, adjustment to various composition, models, and ontologies, and worth, which portrays the expense and worth of big data. Volume describes how large and fast moving big data is. Variety describes how diverse big data is. Big data mining instruments and methods, also known as huge data analytics, have been developed as a result of big data. Big data analytics is the term for a mix of well-known tools and techniques, such as data mining and machine learning that may be used to leverage the important information that is typically secret in big data and produce a connection point as direct and visual analytics.

VII. RECOMMENDATIONS

Future work can be planned for the sharing of speech and video big data, which can be gleaned from a variety of applications. The suggested method can also be expanded to enhance data communication security. With less computational overhead, big data connectivity and information sharing may improve in the future. A strong early admonition administrative instrument in view of big data analytics, machine learning, and artificial intelligence will be developed in the future by identifying more factors that could potentially cause bad bank execution and consolidating these into our models.

REFERENCES

[1] Abdul Majeed, "Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data", Journal of King Saud University - Computer and Information Sciences, Vol.31, Issue 4, pp.426-435, October 2019.

[2] Dash, Tirtharaj and Tanistha Nayak, English Character Recognition using Artificial Neural Network. arXiv preprint arXiv:1306.4621, 2013.

[3] HASHEM, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. In Information Systems. https://doi. org/10.1016/j.is.2014.07.006.

[4] Karn RR, Kudva P, Elfadel IA, "Dynamic auto selection and autotuning of machine learning models for cloud network analytics", IEEE Transaction on Parallel and Distributed Systems, Vol.30(5), pp.1s052–1064, 2019.

[5] KOBIELUS, J., (2018).Deploying Big Data Analytics Applications to the Cloud: Roadmap for Success. Cloud Standards Customer Council.

[6] Pandey, E. ., & Kumar, S. . (2023). Exploring the Generalization Capacity of Over-Parameterized Networks. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 97–112. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2482.

[7] L. Cao, Data science: a comprehensive overview, ACM Computing Surveys (CSUR) 50 (3) (2017) 1–42.

[8] M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science 349 (6245) (2015) 255–260.

[9] MENZES, F.S.D., Liska, G.R., Cirillo, M.A. and Vivanco, M.J.F. (2016) Data Classification with Binary Response through the Boosting Algorithm and Logistic Regression. Expert Systems with Applications, 69, 62-73. https://doi.org/10.1016/ j.eswa.2016.08.014

[10] Kevin Harris, Lee Green, Juan Garcia, Juan Castro, Juan González. Intelligent Personal Assistants in Education: Applications and Challenges. Kuwait Journal of Machine Learning, 2(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/185.

[11] Mobilio M, Orrù M, Riganelli O, Tundo A, Mariani L., "Anomaly detection as-a-service", In: 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW). IEEE; 2019. p. 193–9.

[12] Moustafa N, Creech G, Sitnikova E, Keshk M., "Collaborative anomaly detection framework for handling big data of cloud computing", In: 2017 military

communications and information systems conference (MilCIS). IEEE; 2017. p. 1– 6.

[13] Juan Lopez, Machine Learning-based Recommender Systems for E-commerce , Machine Learning Applications Conference Proceedings, Vol 2 2022.

[14] S. Abd Razak, N. H. Mohd Nazari and A. Al-Dhaqm, "Data Anonymization Using Pseudonym System to Preserve Data Privacy", IEEE Access, Vol. 8, pp.43256-43264, 2020.

[15] S. Sagiroglu and D. Sinanc, "Big data: A review", International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, pp.42-47, 2013.

[16] SITI Nurul Mahfuzah, M., Sazilah, S., & Norasiken, B. (2017). An Analysis of Gamification Elements in Online Learning to Enhance Learning Engagement. 6th International Conference on Computing & Informatics.

[17] Steger, Carsten, Markus Ulrich, and Christian Wiedemann, Machine vision algorithms and applications. 2018: John Wiley & Sons.

[18] Stephen Kaisler, Frank Armour, J. Alberto Espinosa, William Money, “Big Data: Issues and Challenges Moving Forward", 46th Hawaii International Conference on System Sciences, 2013.

[19] Tsai C, Hsu Y, Lin C, LinW, "Intrusion detection by machine learning: a review", Expert Systems with Applications, Vol.36(10), pp.11994– 12000, 2009.

[20] Sherje, D. N. . (2021). Content Based Image Retrieval Based on Feature Extraction and Classification Using Deep Learning Techniques. Research Journal of Computer Systems and Engineering, 2(1), 16:22. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/14.

[21] V. Dhar, Data science and prediction, Communications of the ACM 56 (12) (2013) 64–73.

[22] White, David, James D Dunn, Alexandra C Schmid, and Richard I Kemp, Error rates in users of automatic face recognition software. PloS one, 2015. 10(10): p. e0139827.

[23] Wilson, B. M. R., Khazaei, B., & Hirsch, L. (2015, November). Enablers and barriers of cloud adoption among Small and Medium Enterprises in Tamil Nadu. In: 2015 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM) (pp. 140-145). IEEE.

[24] Zhang J, “Anomaly detecting and ranking of the cloud computing platform by multi-view learning”, Multimedia Tools Appl. 2019;78:30923–42.