Machine Learning Applied to Health Information Exchange

Filipe Miranda, University of Minho, Portugal

Ana Regina Sousa, University of Minho, Portugal
  https://orcid.org/0000-0002-2988-196X

Julio Duarte, University of Minho, Portugal
  https://orcid.org/0000-0002-5458-3390

António Carlos Abelha, University of Minho, Portugal

José Machado, University of Minho, Portugal*
  https://orcid.org/0000-0003-4121-6169

ABSTRACT

The interest in artificial intelligence (AI) has grown in the last few years. The healthcare community is no exception. The present work is focused on the exchange of medical information, using the Health Level Seven (HL7) international standards. The main objective of the present work is to develop an AI model capable of inferring if for a given hour exists a peak in the number of exchanged messages. To accomplish that, two different deep learning models were created, an artificial neural networks (ANN) and long short-term memory (LSTM). The intention is to observe which is capable to perceive the situation better considering the environment and features of a healthcare facility. Using laboratory-generated data, it was possible to simulate variations and differences in "traffic." Comparing the LSTM vs. ANN model, the first is capable of outputting peaks better but for considered mean values that do not happen. For the given context, predicting the peak is essential, so the LSTM is the right choice and uses fewer features that are good regarding performance.

KEYWORDS

Artificial Intelligence, Artificial Neural Networks, Health Level Seven, Information Exchange, Long Short-Term Memory, Machine Learning

INTRODUCTION

Machine learning is a branch of Artificial Intelligence (AI) that constructs self-learning algorithms capable of labeling new data after extracting knowledge from previous examples. Some authors considered the evolution of pattern recognition as one of the main topics nowadays (Wooldridge & Jennings, 1995). Due to extreme importance and data availability, the healthcare community focuses on models that help predict diagnostics, create decision support systems (predict treatment outcome),

and increase medical image quality. New and more complex models are used to demand knowledge representation from supervised to unsupervised and reinforced learning. In the present case study, the Artificial Neural Networks (ANN) are nonlinear statistical learning algorithms inspired by the neurons' structure and functionality.

In recent years the increase in computational capacity and data availability brought to the table new and innovative ways to implement and abstract the complex mathematical operations behind machine learning. The TensorFlow (*Introdução Ao TensorFlow*, n.d.) library and KERAS (Sebastian & Vahid, 2017) are among the top APIs used worldwide. Due to data availability and vital importance, healthcare is receiving attention from machine learning experts and developers worldwide, mainly on treatment predictions, outcomes, treatment guidelines, and without a doubt, medical image processing. No work related to the usage of a model applied to HL7 messages exchanged was identified.

Many healthcare organizations worldwide use Health Level Seven (HL7) standardized messages to exchange information between devices that perform different tasks and "talk" other languages inside an institution. For example, a physician orders a blood exam. That request represents a trigger event that leads to a message creation (request) following HL7 v2 standards. The interface is responsible for intermediating the exchange of data, mainly using sockets. That standard is known as pipe type "|" once every segment on a field is separated by a pipe. HL7 standards constitute the backbone for the healthcare organization's interoperability. Thus, that kind of system must work 24/7 without failures. So, every system that monitors and prevents errors on those interfaces is vital. After research, a gap was identified among the existing HL7 architectures. No ML modules were identified regarding the error prediction or stress hours identification directly related to HL7 interfaces. On the other hand, some works on diagnosis and patient's medical history were found regarding semantic interoperability.

To clarify the subject, interoperability in healthcare is the capability for two or more systems to exchange data and ultimately generate information from that (Garde et al., 2007). Without international standards, it is not possible to accomplish semantic and technical interoperability. HL7 International, OpenEHR, or even SNOMED CT are worldwide recognized organizations for proposing internationally accepted standards (Kalra, 2006). Those standards are used in clinical situations representation, information exchange, or archive/data generation interfaces (*Clinical Knowledge Manager (CKM)*, n.d.). The exchange of information under the HL7 format is essential to a healthcare organization's day-to-day life. Almost all systems are connected using that kind of message, and the failure of one of such web interfaces could cause the complete breakdown of a healthcare institution. Exams are requested, and reports are printed using HL7 as a way to transport the message.

Developing such interfaces for interoperability is complex. Almost perfection and availability are some critical aspects of these interfaces. The other significant part of that specific interface is availability—some seasons with intense work like flu season or different particular situations. The idea is obvious. The healthcare systems do not have a margin for error or misunderstandings.

The present case study intends to infer the possibility of introducing an ANN or LSTM network to predict the number of messages being exchanged under the HL7 format for a given hour inside a healthcare organization. Then that model can be used by other systems to prevent errors and take measures in busy hours. The daily schedule for HL7 agents inside an organization can be approached based on that model as a prediction in a case similar to road traffic on cars or web applications usage.

The paper is organized as follows, the introduction to the subject, background presentation topics, and related works. Mainly AI, NN, can measure performance mechanisms and, finally, interoperability. A section for case study presentation, results, discussion section, and conclusion follows.

BACKGROUND

Artificial Intelligence

Artificial Intelligence is the capability of a computer or computer-controlled robot to perform a task usually developed by intelligent beings (Luh et al., 2019). As John McCarthy defined "…the science

and engineering of making intelligent machines…" (Andresen, 2002). Machine Learning (ML) is a branch of AI. That said, all ML algorithms are AI, but not all AI is ML. For example, symbolic logic (rules engine, knowledge graphs) is AI but not ML. In 1959, Arthur Samuel defined ML as "…field of study that gives computers the ability to learn without being explicitly programmed…" (Samuel, 1959). The last part of the sentence is perhaps the most important ("...Without being explicitly programmed..."), indicating that the algorithm must evolve and adjust itself to respond to new data. Learning is nothing more than optimizing the function on a specific dimension, minimizing the error or increasing the probability of presenting a correct prediction (Samuel, 1959). Supervised, unsupervised, or reinforced constitute the three main ML categories. The first case uses labeled data, direct feedback to predict based on future data. The second finds hidden data structures using data without labels or feedback. The reinforced learning implies a decision process, reward system, and learn through a series of actions (Sebastian & Vahid, 2017).

ML concept has more than 50 years. However, only the increase of computational capabilities and data availability allowed the idea to evolve from theory to practice in real-world scenarios (Luh et al., 2019). Nowadays, all society areas try to implement machine learning to label or find patterns in vast amounts of data (data sets). Healthcare is no exception. A field with a massive amount of structured and unstructured data. More important than that exists the never-ending desire to give the best possible care services. Luh et al. proposed in (Luh et al., 2019) a review document considering how AI can impact the radiation's oncology. The author pointed out works on toxicity prediction, possible outcome prediction, medical imaging enchainment, predicted treatment response, treatment planning, or even working on patient safety, regarding the outcome of a treatment prediction. A machine learning model estimated with better accuracy the patients' life expectancy than most physicians that errored 22% of the time.

Regarding treatment planning, the identification of tumor mass volumes is vital. Deep learning techniques are used to delineate targets on a voxel-by-voxel basis. Through the training of a deep learning convolutional neural network with data from 52 different patient's treatment plans allowed a voxel-based classification. That constituted the work of Cardenas and colleagues in (Cardenas et al., 2018).

In "Bending the Artificial Intelligence Curve for Radiology: Informatics Tools from ACR TO RSNA" (Kohli et al., 2019), presented a series of use cases that introduces a series of informatics tools and strategies to improve vendor adoption and standard creation. As well as possible ways to organize and connect systems inside a radiology department to enable AI applications and decision support systems. The following section will indulge in the main concepts of Artificial Neural Networks and explain why they are so used nowadays.
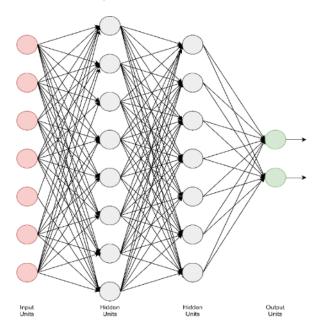
Artificial Neural Networks

Neural networks (NN) are algorithms modeled by the human brain composed of neurons, so NN uses artificial ones. A natural neuron works using the accumulation of electric potential, which causes a discharge across the axon when exceeding a value. It is then possible to replicate that to the artificial neurons (Cardenas et al., 2018). Thousands of neurons are called units arranged in layers, each connected to the layers on either side. Figure 1 represents the overall architecture of an ANN, the input layers receiving data from the outside, and the hidden layers that constitute the central part of the "brain". That's an example of a fully connected ANN once each unit is connected to all the units in the next layer. Usually represented with a line is the weight, positive or negative, that acts as an exciting or repressive element of the next unit. Higher the weight, more influence has on subsequent. The same happens with the brain cells triggering one another across the synapses. So, the layer applies a transformation to the entering data based on the parameterized by its weights (Sebastian & Vahid, 2017).

Overall learning means finding the weights that enable the correct mapping of the inputs to the outputs, minimizing the error when input is labeled. That reduction of the error means, in every

Figure 1. Artificial Neural Networkstructure example

iteration, finding which will be the next direction of the function that will significantly impact error reduction. To accomplish that without getting stuck on local minimums or saddle points, gradient descent or stochastic gradient descent are among the most used. To control the output and calculate the loss score, a loss or objective function computes the distance between the output and the actual output during the training stage. The ANN model uses that value as a feedback signal to the network through the backpropagation algorithm (Nikhil Ketkar, 2017). The initial weights are random values; the output will be far from the truth. With every example, the backpropagation adjusts the weights in the right direction, and the loss value decreases.

Two scenarios are possible regarding the output layer; quantitative variables using regressions or categorical (Yes or No) known as a classification problem. The hypermeters defined before the test have a direct impact on the overall results follows an enumeration of some of them (Nikhil Ketkar, 2017):

- Weight initial values, generally a uniform distribution.
- The number of hidden layers, a small number, could lead to underfitting.
- Dropout is used to avoid overfitting, increasing generalizing power.
- The activation function introduces nonlinearity to models and allows the models to learn nonlinear prediction boundaries. Rectifier activation function, sigmoid (binary prediction), and SoftMax (multi-class prediction) are among the most popular.
- Learning rate defines how quickly a network updates its parameters, low and converges smoothly, to large and may not converge at all.
- Momentum, based on previous steps, helps prevent oscillations generally between 0.5 and 0.9.
- The number of epochs and the number of the data set is shown to the ANN while training; the periods should stop growing when accuracy in the validation set starts decreasing. However, the accuracy of training still increases (overfitting).
- The batch size was introduced to prevent gradient descent from getting stuck at the local minimum of the loss function, thinking it is a general minimum. The possible values are 32,64,128,256, and so on.

Figure 2 illustrates the flow inside two ANN. The sum of all inputs is then passed through an activation function. Next, some equations of the flow inside an ANN will be detailed. Equation 1 merely represents what happens inside a neuron. The x is the feature, B_1 represents the slope parameter, clearer how much changes xB_0 represents the bias term, and the $sigmoid$ activation function is used to output the final probability:
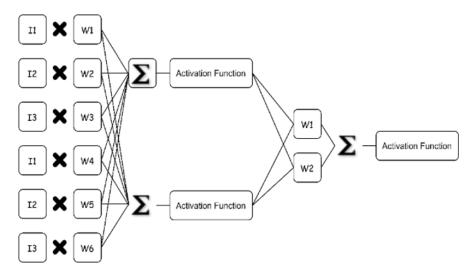
$$sigmoid\left(B_1 * x + B_0\right) = P \cdot Output \tag{1}$$

Changing the weight has a reverberating effect across all the other neurons and activation in the subsequent layers because a neuron has its little model (Nikhil Ketkar, 2017). Find the weights and biases that minimize the loss function (which represents how wrong are the predictions). To use gradient descent, it is essential to know the gradient of the loss function, the vector that points in the direction of most significant steepness and takes steps in the opposite direction. Enabling that backpropagation, the weights and biases across the entire network are updated. Equation 2 tries to summarize the stochastic mini-batch process that avoids local minimums and possible saddle points for the difference in weights for given i j connection over the time k:

$$\Delta w_{ij} = -\Sigma \epsilon\, y_i^{(k)} y_j^{(k)} \left(1 - y_j^{(k)}\right) \frac{\partial E^{(k)}}{\partial y_j^{(k)}}_{k\epsilon\ minibatch} \tag{2}$$

The scientific community embraced the implementation of neural networks in the most varied contexts. Compiling the most exciting works from 2013 to 2018, Maisa Doud and Michael Mayo presented in (Daoud & Mayo, 2019) a list of neural network-based models for cancer prediction. The authors pointed out that ANN models are used as filters, predictors, and clustering methods. Simultaneously, the major problems common to the analyzed papers were developing a consistent architecture considering the different inputs. The authors concluded that the role of the neural network determines its general architecture.

Figure 2. A2 neuron ANN from input to output Keras and TensorFlow

In "Diagnosing internal illnesses using pervasive healthcare computing and neural networks" (Bayraktar et al., 2011), Canan Bayraktar et al. presented a distributed healthcare system that uses ANNs to compute diagnostic prediction. The paper has a detailed representation of the 103 features used to calculate results. The author used blood report values as input. The output layer with 19 nodes has a list of possible diagnoses. Although brief and without graphical contextual error values and accuracy, the paper is exciting and pertinent. Md. Shamim Talukder et al. worked with SEM-Neural Networks to predict wearable technology acceptance in advanced age personas (Talukder et al., 2020). With seven features and one output node, the model was implemented with success and enabled interesting conclusions once the author started to change the input values.

Not healthcare-related, Victor F. Pallares and colleagues presented in (Fernandez Pallarés et al., 2019) a possible solution to traffic management considering the electric car's paradigm. The team purposes an architecture to power supply availability. The traffic prediction comes in hand to prepare the route vehicle, which is used as input to the model a mathematical formulation of city traffic. The input from the series is used as a data generator for the ANN to find patterns and reach conclusions. Once the purpose of the present work is to find traffic zones in the exchange of healthcare information, it's pertinent to look at other traffic perdition works. The following section will introduce the Long Short-Term Memory kind of networks that are best suited for temporal series analysis. That kind retains information from previous steps to influence the current step.

Long Short-Term Memory

A recurrent neural network (RNN), first introduced in 1980, is different from the so-called feed-forward because it can maintain a state between uses of the network (Nikhil Ketkar, 2017). An ANN passes to the next neuron a computed activation of the current step. Au contraire the RNN preserves the activation from the previous time step accumulating the status of the network. Of course, adjusting weights based on error derivatives needs a facelift, so now they consider averaging or summing all the error derivatives over all the connections within the same set. But a problem arises if the weights of the model are small initially, the derivatives will inch even further near zero after some runs. So the conclusion is clear; the problem of vanishing gradients prevents the model from learning long-time relationships between data (Nikhil Ketkar, 2017).

In 1991 Sepp Hochreiter and colleagues proposed the Short-Term Long Memory (LSTM) (Hochreiter & Schmidhuber, 1997). The model solves the upper mentioned problem by combining a RNN with a gradient-based algorithm. The output of a cell still has the transformation from the last iteration but suffers from many subsequent ones. To reliably transmit important information, several steps into the future are required. During an approach organized in gates, the system uses a hidden layer called the memory cell.

The output of the current step is represented by $C^{(t)}$ whereas the previous one is $C^{(t-1)}$. The information flows through gates (forget, input, output), containing sigmoid activation functions and hyperbolic tangent (tangent), element-wise product, an element-wise summation that transforms the input $x^{(t)}$ (Nikhil Ketkar, 2017). Equation 3 represents the hidden units at a given time step (indicates an element-wise product). That is the result of the forget gate, the input gate, the input node:

$$h^{(t)} = O_t \odot tanh\left(C^{(t)}\right) \tag{3}$$

The output gate will decide the way hidden units will suffer the update stated in equation number 4:

$$O_t = \sigma\left(W_{xo}x^{(t)} + W_{ho}h^{(t-1)} + b_0\right) \tag{4}$$

The usage of a sigmoid function decides which values to let through as a 0 or 1 system. Whereas the $tanh$ function provides the weightage of the values passed stamped their level of importance from -1 to 1. Both equations were derived from the book Deep Learning with Python (Nikhil Ketkar, 2017) and serve as examples of what happens inside an LSTM model.

The LSTM is very useful in time series forecasts like the stock market and sales forecast using the previous hour or previous month. Another area where LSTM is used is in sentiment analysis from text or even textual translation.

In the field of acoustics Fatih Ertham (Ertam, 2019) presented a paper to predict the person's gender from the voice. Using ten features in input and 100 hidden neurons over a 30-epoch training session, the author achieved a 98.4% accuracy. In the same field, Hasim Sak and colleagues (Sak et al., n.d.) introduced two LSTM layers in a row to accomplish large-scale acoustic modeling. The paper served to prove the validity of the LSTM, and the combination of the two layers was essential to increase accuracy. Another critical point of the work was the introduction of the distribution capabilities to reduce training time.

Keras is an open-source neural network framework for Python that allows defining and training of deep-learning models. It also allows the code to run on CPU and GPU, with a simple interface. The API runs on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or even R. TensorFlow is a free, open-source computational framework for dataflow based on a symbolic math library that offers excellent support to machine learning. It provides a very vast offer of toolkits from a lower to a high level when implemented in Python. The critical point is that it is developed by Google Brain and released under Apache License 2.0.

Multi-Agent System

Exchange of healthcare information using HL7 FHIR or V2 of the standard is a significant task serving as the healthcare institutions' backbone. Michael Wooldridge defines an intelligent agent as "…as an autonomous system, capable of interacting with other agents to satisfy its design objectives, is a natural one for software designers…". The same author states that an agent system has the following properties:

- Autonomy, using the agent status and system states, can adjust behavior.
- Reactivity, agents perceive the environment and respond in accordance.
- Proactiveness, agents perform the work, and besides that can take the initiative.
- Social ability, agents interact with other agents, using a common language and usually performing social activities.

Most agents are used on the exchange of medical information under version 2 of the format regarding healthcare. That will be detailed in the next section.

HL7 Interface Engine

The intended model will infer healthcare-related messages generated from an HL7 interface engine. Hl7 International specifies through standards in which format or manner information should be shared across the internet between distinct systems. That connection is called an interface. The development of such algorithms is in charge of the developer. HL7 international only purposes of standards and guidelines. An HL7 interface includes:

- Export endpoint for the sending application.
- Import endpoint for the receiving application.

- A method for moving data between two endpoints (TCP/IP), for example.
- A logging mechanism for prior inference from results.

In 2012 Miguel Miranda et al. proposed a multi-agent system destined to exchange HL7 messages. Using JADE Java frameworks and ontologies, the work was an exciting approach to the thematic. Some years passed, and it is essential to discover new paths to the exchange of medical information. The following section will present the present case study in detail.

CASE STUDY PRESENTATION

The introduction of a machine learning model requires a set of subtasks before implementation or validation. Those can be detailed as definitions of the problem, data preparation, algorithm investigation and implementation, improving results, and finally presenting results to other subsystems or exciting parts. The present model has intended for the "hot hours" prediction of healthcare HL7 v2 messages on a healthcare organization based on previous messages.

In a nutshell, the case study tries to find the answer to the following research questions:

- What are Deep Learning and possible application techniques?
- What features define the exchange of healthcare information between different systems using the local intranet as transportation?
- How to adapt a multi-agent system architecture to take advantage of the ML module?
- Which Deep Learning model best suits the exchange of HL7 V2 messages considering a time-series analysis?
- What were the implementation impacts and possible future changes?
- What were the lessons learned from this simple implementation?

To evaluate the results of the data model, Skewness and Kurtosis values were measured. The loss value was the barometer used to assess the models. Also, considering the scenario, peak detection was an important aspect. The loss value was retrieved by comparing the predicted values with the ones separated from the dataset as a validation dataset.

As stated, two different models were implemented, one standard artificial neural network and a second recurrent one using the LSTM methodology. The reason is the capability of these models to deal with time series and time connections. The difference is that ANN will include all the features demonstrated in Table 1 au contraire the LSTM model will feature an hourly update off message number and use only message number and date. The same approach as a stock market prediction.

The construction of the ANN and LSTM model using Keras and TensorFlow was conceived using Sequential type. All use a Sigmoid output layer to consider a value between 0 and 1 that is converted to an actual message number. The training stage's essential point was to prevent overfitting and underfitting (train too much or too little to fit the training set too well). In both scenarios, a dropout of 0.3 and a batch size of 32 and 10 were used.

The ANN and LSTM models were created, trained, tested, and finally validated using that data. The following section will answer some of the questions.
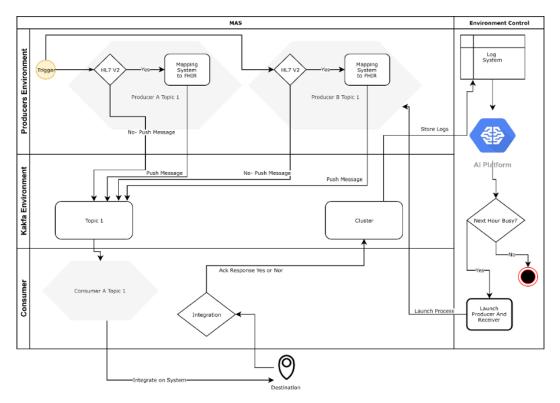
RESULTS

To develop the reason model, as stated on the right-side Figure 3, a data set following the subject was assembled and generated in the laboratory once that kind of data set was not available. To mimic reality, some points were considered, the flu season, the summer break, and weekdays. Those properties will increase or decrease the number of messages being exchanged. At the beginning of the week

Table 1. Model implemented based on data set properties

ID	Attribute	Units
1	Date	YYYY-MM-DD HH24-mi-SS
2	Week of the Year	[1-53]
3	Hour of the Day	[0-24]
4	Day of week	[0-7]
5	Hour of the Day	[0-24]
6	Flu Season	[0-1]
7	Begin of the Week	[0-1]
8	Hot Hour	[0-1]
9	Summer Break	[0-1]
10	Month	[1-12]
11	Number Hospitalizations	[0-500]
12	Number Emergencies Hour Before	[0-500]
13	Laboratory Hour	[0-1]
14	MCDT Hour	[0-1]
15	Number of messages	[0-1000000]

Figure 3. Multi-Agent System proposed architecture. Be aware that the present case study's work focuses on the right part of the image. The rest of the image is broad research and development work.

and in flu season, the number of people in the emergency room will be higher, so more messages are exchanged. Influencing the number of messages is the hours during the day where the healthcare facility is open for external exams like blood analysis of x-rays. Those will impact the HL7 message's numbers; once every time a patient confirms his/her presence, the PACS system will be notified to create or update the worklist. That increases traffic during certain hours. After creation, the dataset was tested for validity and applicability based on a set of properties by literature.

Figure 3 serves as an answer to research question number 3 and how a model like this can be integrated with an already existing system.

With data from 2010 to 2019, the data was prepared for the training stage, which means that labels or message numbers were separated in the different dataset; all values were normalized on a scale of 0 to 1. The primary three separated data sets were created, the training one, the test one, and the validation one. Off-course the LSTM, as explained, needs more preparation once only the hour-of-day is required and message numbers of training the model. Considering a lookback variable of 24 hours, an array with the last 24 outcomes is created with an output of the next hour. On the next time step, the t will feature the t-24 outcomes as well.

Regarding the model definition, the previous section presents the essential aspects of the dataset. Matias Kraus and colleagues explained in (Kraus et al., 2020) various finance and sales forecast models. The last model presented demonstrates a temporal series close to the one used in the present case study. Figure 3 represents the proposed architecture compromising machine learning and information exchange in format HL7 FHIR that is the basis for responding to question number 3. However, the image contains a lot more elements than the ones used in the present work. From the image, it is essential to retain the right end side.

The complete information of the dataset is available in Figure 4. Regarding the test for data normality, it checks if the data follows the normal distribution. The skewness is a measure of the asymmetry of the probability distribution of a random variable about its means. The kurtosis tells the height and sharpness of the central peak regarding a standard bell curve. The results were for kurtosis of normal distribution GoodData Documentation (Kurtosis, n.d.): 2.390 and skewness of normal distribution (Kurtosis, n.d.): -1.1404. That information is available in Figure 5 on the left-hand side of the graph.

Using the previous data and model, the LSTM and the ANN models were constructed. In the first case, the information of the model is as follows:
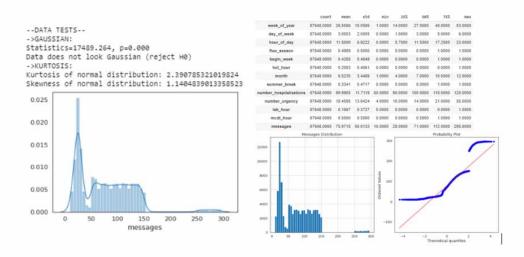
- Input layer with 100 LSTM nodes;
- A dropout of 0.3 between layers to prevent overfitting;
- Three hidden layers with 100 nodes "ReLu" activated;
- One output layer "sigmoid" activated node;
- Loss function "mean_squared_error";
- Optimizer "adam";
- Metrics "mean_absulote_error";
- Epochs 20;
- Batch size 32.

The ANN model has the following characteristics:

- Input layer with 12 nodes "ReLu" activated;
- A dropout of 0.3 between layers to prevent overfitting;
- 4 hidden layers with a total of 220 nodes "ReLu" activated;
- 1 output layer "sigmoid" activated node;
- Optimizer "adam";
- Loss function "mean_squared_error";

Figure 4. Dataset-related information. The top board represents a sample of the dataset used. And essential information for rows and columns. The bottom row on the left side has the Kurtosis and Skewness results and distribution of message numbers. Then the right end side, more detailed information representing the mean and std deviation of the dataset.

- Metrics "mean_absulote_error";
- Epochs 20;
- Batch size 10.

The output is a dense layer with just one node that prompts the number of messages.

The following figures will demonstrate the results of the previously detailed modes. All the figures will feature the training loss vs test loss—the final chart features the model's actual vs predicted values.

The results of the ANN model.

LSTM results for the model implemented for the same dataset with the nodes' adjustments required.

The following section will introduce the discussion based on the present section results. It was mainly analyzing the charts and perceiving the best option to follow in the future.

DISCUSSION

Research question number 1 already was answered with the introduction and background. It is essential to notice that deep learning is complex because a tradeoff between hyperparameters needs to balance the results and prevent over and underfitting. The application techniques are immense, every one of them with particular considerations. For example, to implement an ANN, we consider a broad dataset, but for the LSTM scenario, a fraction of the columns are considered. A lot of examples were considered during the background and introduction sections. It is essential to highlight that the

model definition is perhaps more important than the parameters definition stage once representing the machine's reality.

Table 1, with the set of possible features that define the problem in hands, is the ground to answer research question number 2. From that list, some features were dropped to increase the model results because we're introducing redundancy and lowering the results' accuracy. For example, the month and week of the year only cause noise. Figure 4, the top side, gives an insight into the data set with all columns. The hour is essential and other conditions that impact the affluence of persons into the emergency room. The total of internments until that specific hour also affects the number of messages once a physician can order blood exams or examinations on that patient. That value was derived per hospital location and season. Also related to the model, Skewness and Kurtosis values are both positive but higher than 1. The skewness case reveals that the distribution is highly skewed, meaning a divergence from the normal distribution and the symmetry. The fact that kurtosis is more significant than 1 means that a tail of the distribution identifies that the model has some outliers but still on an acceptable level. Figure 4 also has insights into the data to give a general overview of the model data quality.

Regarding the results of the ANN model, the loss and absolute error presented good values. Both less than 0.1, as Figure 5 and 6 represent. It is possible to perceive that decrease over the epochs. We can sense that predictions from Figure 7 can follow the peaks of messages and some of the lows. Still, some underfitting was found for the mean value, perhaps motivated by an excess of correlated values on the model or 20 epochs was a little less to the complexity of the model. Regarding hidden units, the model already has more than 500, which is a high number. The data set is extensive, as proved before.

The LSTM model uses the same LSTM (input layer and three hidden) nodes and Dense "ReLu" and "Sigmoid" as activated as well for the last two layers. Figure 9 represents the loss's values for three lookback values considered, 6 12, and 24 hours. From the figure, we can perceive that the best values are for 24 hours. So, the model learns better with more data considered as input. That is expected because the reason units have more values to consider. Looking at Figure 8, we get a sense of the prediction using the validation set.

Interestingly, the 24-hour case has worse results than the other with underfitting in the chart's middle. Both 6 and 12 can understand the intermediate values a little better. They also perceive the peaks better than the 24-hour case. Regarding performance, the 6 hours is faster on the training stage. That's why we can use a 6-hour time window to output results for a given hour.
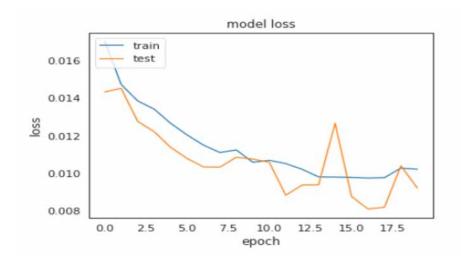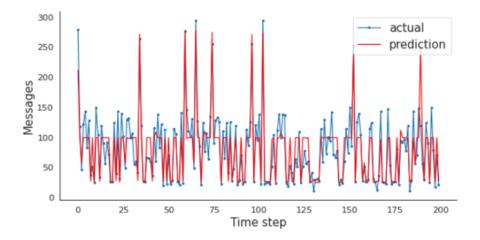
Figure 5. ANN model loss (Train vs. Test)

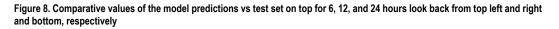**Figure 6. Model absolute error for train vs test on the ANN model**

Figure 7. Comparative values of the model predictions vs test set for the ANN model

The better loss values are from the model ANN, although that other one is not far away. The primary concern is the capacity of the LSTM to perceive the peaks and reach conclusions. The ANN model's primary concern is the overfitting with the trained model that prepares the model to train well but does not happen in reality. It is essential to point out that only 60% of the data set was used in the model training stage. The others are used for testing and later verification of the model.

CONCLUSION

The research questions presented at the beginning of the present case study were answered through the last two sections; when creating a multi-agent system capable of integrating an AI model, some points must be addressed. The capacity to keep logs for future training processes. How to proceed with that training regarding that for future predictions, the data set must be updated. Another point more technical is how to offer the forecast to the MAS, a service where the number of messages will serve as a measure to activate second measures or label the past hours as busy or free and only output that value. We could also use the decision to take actions and train the network for that job. That is a

Figure 8. Comparative values of the model predictions vs test set on top for 6, 12, and 24 hours look back from top left and right and bottom, respectively
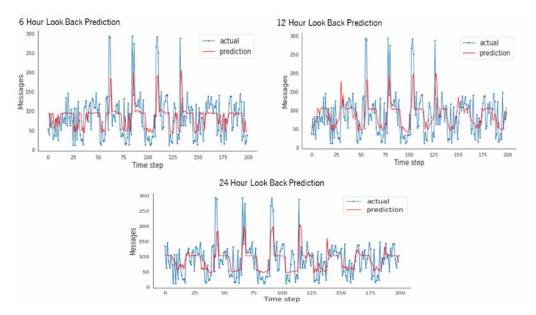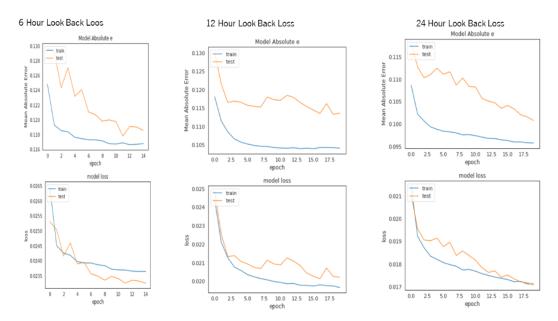

Figure 9. Case study 5 - LSTM absolute and model loss (train vs test). Regarding the loss values from left to right representing 6, 12, and 24 hours look back.

future measure. The present study focused only on predicting the message number to the next hour and acted upon that without reasoning.

Regarding the lessons learned to the future and how they can influence the present work, it is evident that both models can be improved in the future. The data used was not optimal, and also, the number of layers perhaps introduced overfitting of the model. Another point is the redundancy of the

features used. If we pay attention to them, all are related to time definition, year and month influence almost every parameter, and the influence of the number of messages. That introduces redundancy and increases the change of overfitting of the model. Perhaps reducing the number of features and improving the specificity presents the best results.

Another central point is the duality between the model using LSTM or regular ANN. Although the latter could present better accuracy, the loss in all the training was more significant. In that case, the best model is the LSTM. Another central point is the usage of prior values to compute the new ones. And the impact that has on the system. With the LSTM, an iterative process can calculate the value and use it in the following hour's input. Using the ANN, the loose values are computed one at a time without connections or correlation. It would be challenging to calculate the temperature, number of emergencies. Perhaps three different feeding algorithms to compute the messages would be necessary. Using only the message component is simpler without losing consistency.

NOTE

No personal data was used for the development of the present work.

REFERENCES

Andresen, S. L. (2002). John McCarthy: Father of AI. In IEEE Intelligent Systems (Vol. 17, Issue 5, pp. 84–85). doi:10.1109/MIS.2002.1039837

Bayraktar, C., Karan, O., & Gümüşkaya, H. (2011). Diagnosing internal illnesses using pervasive healthcare computing and neural networks. *Procedia Computer Science*, *3*, 584–588. doi:10.1016/j.procs.2010.12.097

Cardenas, C. E., McCarroll, R. E., Court, L. E., Elgohari, B. A., Elhalawani, H., Fuller, C. D., Kamal, M. J., Meheissen, M. A. M., Mohamed, A. S. R., Rao, A., Williams, B., Wong, A., Yang, J., & Aristophanous, M. (2018). Deep Learning Algorithm for Auto-Delineation of High-Risk Oropharyngeal Clinical Target Volumes With Built-In Dice Similarity Coefficient Parameter Optimization Function. *International Journal of Radiation Oncology, Biology, Physics*, *101*(2), 468–478. doi:10.1016/j.ijrobp.2018.01.114 PMID:29559291

Clinical Knowledge Manager (CKM). (n.d.). Retrieved November 26, 2020, from https://ckm.openehr.org/ckm/retrieveResources?list=true

Daoud, M., & Mayo, M. (2019). A survey of neural network-based cancer prediction models from microarray data. In *Artificial Intelligence in Medicine* (Vol. 97, pp. 204–214). Elsevier BV., doi:10.1016/j.artmed.2019.01.006

Ertam, F. (2019). An effective gender recognition approach using voice data via deeper LSTM networks. *Applied Acoustics*, *156*, 351–358. doi:10.1016/j.apacoust.2019.07.033

Fernandez Pallarés, V., Cebollada, J. C. G., & Martínez, A. R. (2019). Interoperability network model for traffic forecast and full electric vehicles power supply management within the smart city. *Ad Hoc Networks*, *93*, 101929. doi:10.1016/j.adhoc.2019.101929

Garde, S., Knaup, P., Hovenga, E. J. S., & Heard, S. (2007). Towards Semantic Interoperability for Electronic Health Records Domain Knowledge Governance for openEHR Archetypes. *MethodsInf*, *46*(03), 332–343. doi:10.1160/ME5001 PMID:17492120

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735 PMID:9377276

Introdução ao TensorFlow. (n.d.). Retrieved November 26, 2020, from https://www.tensorflow.org/learn

Kalra, D. (2006). Electronic health record standards. In Yearbook of medical informatics (pp. 136–144). doi:10.1055/s-0038-1638463

Kohli, M., Alkasab, T., Wang, K., Heilbrun, M. E., Flanders, A. E., Dreyer, K., & Kahn, C. E. Jr. (2019). Bending the Artificial Intelligence Curve for Radiology: Informatics Tools From ACR and RSNA. *Journal of the American College of Radiology*, *16*(10), 1464–1470. doi:10.1016/j.jacr.2019.06.009 PMID:31319078

Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*, *281*(3), 628–641. doi:10.1016/j.ejor.2019.09.018

Kurtosis, S. (n.d.). *Normality Testing*. Retrieved November 26, 2020, from https://help.gooddata.com/doc/en/reporting-and-dashboards/ maql-analytical-query-language/ maql-expression-reference/ aggregation-functions/ statistical-functions/ predictive-statistical-use-cases/ normality-testing-skewness-and-kurtosis

Luh, J. Y., Thompson, R. F., & Lin, S. (2019). Clinical Documentation and Patient Care Using Artificial Intelligence in Radiation Oncology. *Journal of the American College of Radiology*, *16*(9), 1343–1346. doi:10.1016/j.jacr.2019.05.044 PMID:31238022

Nikhil Ketkar. (2017). *Deep Learning with Python A Hands-on Introduction*. 10.1007/978-1-4842-2766-4

Sak, H. H., Senior, A., & Google, B. (n.d.). *Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling*. Academic Press.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. doi:10.1147/rd.33.0210

Sebastian, R., & Vahid, M. (2017). *Python Machine Learning*. Packt Publishing Ltd.

Talukder, M. S., Sorwar, G., Bao, Y., Ahmed, J. U., & Palash, M. A. S. (2020). Predicting antecedents of wearable healthcare technology acceptance by elderly: A combined SEM-Neural Network approach. *Technological Forecasting and Social Change*, *150*, 119793. doi:10.1016/j.techfore.2019.119793

Wooldridge, M. J., & Jennings, N. R. (1995). *Intelligent Agents: Theory and Practice*. Cambridge University Press.

Regina Sousa is a Ph.D. student in Biomedical Engineering with technical skills in Artificial Intelligence, Business Intelligence, Big Data, Healthcare Information Systems, among others. She collaborated in a professional project with TecMinho in the years 2017 and 2018 called DermoID -Núcleo de Investigação e Desenvolvimento de Dermocosméticos. In 2018 she was a trainer of children's computer programming in the company HappyCode Braga. In 2019 she joined geoatributo as a software analyst where she prevailed until the end of the year. She took leave from the profession to start her Ph.D. in Biomedical Engineering in the Medical Informatics Branch. In Jan 2020 she won a research fellowship in the Factory of the Future: Smart Facturing project where he still collaborates. In Nov 2020 experienced the first activity as a guest lecturer in the course of Asian Institute of Technology (AIT) - Data Modeling and Management in the School of Engineering and Technology (SET) Asian Institute of Technology (AIT), Thailand. In Jan 2021 she was a lecturer at Instituto Piaget Vila Nova de Gaia in the Information Systems course. So far she has published 2 papers and 6 more are submitted awaiting approval.

Júlio Duarte is an Auxiliary Researcher in Universidade do Minho. He got his PhD in Biomedical Engineering, in 2015. He is a researcher of the ALGORITMI research centre, in the line Computer Science and Technology (CST) and the group Knowledge and Data Engineering (KDE). His research interests span the domain of Biomedical Informatics, Electronic Health Records, Interoperability, Databases, Business Intelligence and Applied Artificial Intelligence.

José Machado is an Associate Professor with Habilitation in the Department of Informatics, Universidade do Minho. He got his PhD in Informatics, in 2002, and Habilitation in 2011. He is a researcher of the ALGORITMI research centre, in the line Computer Science and Technology (CST) and the group Knowledge and Data Engineering (KDE). He is the coordinator of the CST. His research interests span the domain of Biomedical Informatics, Electronic Health Records, Interoperability, Databases, Business Intelligence and Applied Artificial Intelligence. He is the header of the CST since 2015 and the director of the former Computer Science and Technology Center since September 2013.