Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 9s (2025)
https://www.jneonatalsurg.com

Artificial Intelligence, Diagnostic and Prevention Tools to Deal with Covid 19 In Algeria

Adel Abdelhadi*[a], Ouahab Kadri[b], Fateh Merahi[b], Mohamed Taki Eddine Seddik[a]

[a]Department of computer science, University of Batna 2, Batna, Algeria.
Email ID: a.abdelhadi@univ-batna2.dz
[a]Department of computer science, University of Batna 2, Batna, Algeria.
Email ID: m.seddik@univ-batna2.dz
[b]Department of Statistics and data science, University of Batna 2, Batna, Algeria.
Email ID: o.kadri@univ-batna2.dz
[b]Department of Statistics and data science, University of Batna 2, Batna, Algeria.
Email ID: f.merahi@univ-batna2.dz
***Corresponding Author:**
Email ID: abdelhadi@univ-batna2.dz

Cite this paper as: Adel Abdelhadi, Ouahab Kadri, Fateh Merahi, Mohamed Taki Eddine Seddik, (2025) Artificial Intelligence, Diagnostic and Prevention Tools to Deal with Covid 19 In Algeria. *Journal of Neonatal Surgery*, 14 (9s), 1-14.

ABSTRACT

In this work, we suggest an original architecture for medical self-diagnosis. Two intelligent models have been created to interact with patients and suggest appropriate medical treatments. This architecture enhances the obtained results through the integration of a Chatbot and an image classification system. The originality of our work lies in the design and development of novel diagnostic procedures. Another unique aspect of our study is the introduction of an architecture that combines text and image classification to identify suitable medications. To validate our proposal, we used COVID-19 disease as a case study, enabling us to guide patients with the assistance of an intelligent agent. Our model was trained using Convolutional Neural Networks (CNNs) on a dataset comprising three classes: COVID, normal, and viral pneumonia. We introduced a new stopping criterion based on accuracy, effectively reducing learning time and preventing confusion. In the latter part of this paper, we examined the consumption patterns of anticoagulants, antibiotics, anti-inflammatories, and analgesics over two years: pre-COVID-19 (2018-2019) and through COVID-19 (2020-2021). Our dataset was compiled from data collected at Mohamed Boudiaf Barika Hospital in Algeria. Based on this study, we proposed a predictive system for drug consumption trends.

Keywords: *Medical self-diagnosis, COVID-19 disease, CNN, drug consumption, prediction system*

1. INTRODUCTION

During this global health situation, health has become the main concern for the majority of human beings. Unfortunately, at the level of Algeria (North Africa), the limited number of doctors cannot cover the medical needs of our society. This problem has made a large percentage of our population take drugs without consulting a doctor (Chebout and Kabour 2022).

This pandemic does not end because of the appearance of new variants of this virus. Several software solutions have been proposed to reduce the pressure on hospitals and control the consumption of drugs. The majority of proposals addressed only one type of disease. This article presents a simple interactive interface that hides the complexity of processing a database. The main objective of the proposed software is to ensure a diagnosis similar to that of a human doctor using a large base of experiences and cases treated. The high number of smartphones has made it necessary to develop applications suitable for the Android platform. We have integrated an offline mode to our application that can work without the existence of the Internet because of the unavailability of a WIFI connection in most public places. Linguistic diversity forced us to use English. But the mastery of this language is not at the expected level. Therefore, any solution must consider this reality and not propose ambiguous solutions that push the patient to go to the hospital and catch other diseases. A traditional remote diagnosis solution based on discussion forums and involving a doctor's permanent presence to answer patients' questions is not always possible. Another solution is to save all the doctor's knowledge in a file tree which is very difficult to browse and to find the desired information (Tsuru et al. 2022).

Currently, the most common solution is to use chatbots to replace live chat. This system must allow different types of communication and does not require the presence of a doctor (Fan et al. 2021). It is based on natural language processing

(NLP) and image analysis. To be this solution efficient and fast, a good image classification algorithm and another for text classification are necessary. Thus, the goal is to ensure an accurate and short conversation (Anki, Bustamam, and Buyung 2021). The term Covid-19 is used to refer to the Coronavirus disease which appears in 2019. The cause of this disease is the SARS-CoV-2 virus which belongs to the Coronaviridae family. There are several hypotheses about the origin of this disease. The first cases appeared in the city of Wuhan in Hubei province in China at the end of 2019. A few months were enough for this disease to spread around the world. Covid-19 is a dangerous disease, especially in elderly patients who suffer from chronic illnesses. The main way of infection is by breathing the small droplets emitted by the patients. The existence of asymptomatic patients increased the rate of infection. In the emergency department, doctors recommend a series of tests. We cite here some examinations, NFS (to detect anemia or the presence of infections (treatment with antibiotics), thrombocytopenia in the presence of a decrease in the number of platelets (avoid treatment by anticoagulant), Serum electrolytes, kidney function, hepatic check, D-Dimers, (fibrin/coagulation derivatives, for treatment with anticoagulants), LDH, CPK (lactate dehydrogenase: lesion marker), CRP (C-reactive-protein associated with acute inflammation for treatment with antibiotics), and blood cultures if fever (for antibiotic treatment).

Among the best-known treatment is preventive treatment. This type of treatment concerns fragile patients. The second type is the curative treatment which concerns the early and advanced phases of the disease. We also use treatments for other diseases that have symptoms similar to those of Covid-19, such as corticosteroids and oxygen therapy. There is a treatment used as a preventive measure that is indicated for frail people. It includes two categories of drugs which are monoclonal antibodies and antivirals. Evusheld dual monoclonal antibody therapy is highly effective against multiple variants of the virus. It is highly recommended for adults and the elderly. The monoclonal antibody Xevudy is a necessary treatment within 5 days of the first symptoms. Like the previous treatment, it is recommended for the same age group. Antiviral Paxlovid is a treatment that may be recommended if the symptoms have appeared for several days and if the patients are not at risk. In Algeria and according to figure 1., Covid-19 went through four waves, the first in March and June 2020, the second in December 2020, the third in June, July, and August 2021, and the fourth in December 2021 and January 2022.

The remainder of this paper is structured as follows: Section 2 presents an overview of key related works. Section 3 outlines the architecture of the Chatbot, while Section 4 details the AI tools utilized in various components of the software. Sections 5 and 6 present the experimental results of image classification and NLP. A statistical analysis of drug consumption is presented in sections 7 et 8. While the overall results and the conclusion are presented in section 9.

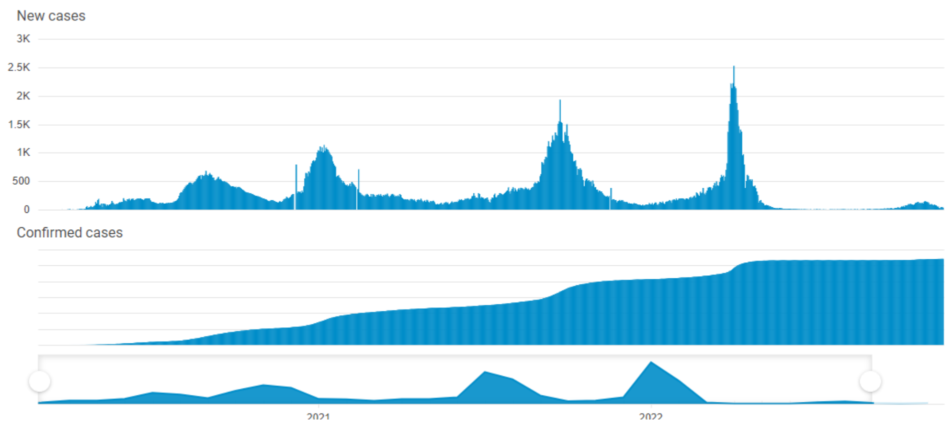

Figure 1. New and confirmed cases in Algeria

2. RELATED WORKS

Recently, machine learning has become a crucial tool for developing chatbots. This section briefly reviews several research studies focused on building medical diagnosis systems using machine learning and deep learning techniques. Manyu and Rajiv introduced an assistant chatbot designed for English-to-English translation, aiming to enhance the model's perplexity and learning rate while optimizing the BLEU score for accurate language translation (Dhyani and Kumar 2021). Temniranrat et al. developed a rice plant disease diagnosis system that operates through a mobile Line application, analyzing images captured in real paddy field environments (Temniranrat et al. 2021). Eslam et al. proposed a smart chatbot system for interacting with users and providing answers related to COVID-19. Their model leverages Google's pre-trained BERT language model and employs a two-phase architecture for question-answering. The first phase utilizes a text classification technique powered by the BERT transformer to categorize text input based on word meanings. The second phase focuses on querying the relevant domain to generate responses. The system was trained and evaluated on Stanford University's SQuAD V2.0, a widely recognized question-answering dataset (Amer et al., 2021).

Several chatbots have been developed to facilitate distance education by leveraging online resources. These academic tools address various challenges in the field. Urquiza-Yllescas et al. proposed a classification framework for educational chatbots based on a systematic mapping study and an iterative examination method. Their work categorizes chatbots, outlines their characteristics and limitations, and explores their potential applications for developers and researchers (Urquiza-Yllescas et al. 2022). Zhang et al. introduced a multi-feature spatial convolutional semantic matching model (BMCSA) that integrates BERT with spatial and semantic feature enrichment. BMCSA utilizes the BERT model to extract semantic features, applies a two-dimensional convolutional network to capture spatial information, and incorporates an Attention mechanism to refine global feature representation. The model's effectiveness was validated using two semantic matching datasets and a text inference dataset (Zhang et al. 2022). Fuwei et al. developed an RNN-based model, Variational Hierarchical Conversation with Subject-Aware Latent Variables (VHCR-T), designed to enhance background information utilization for generating diverse and contextually rich responses. The model comprises three latent variable levels: an aggregate-level latent variable for contextual representation, a topic-level latent variable for capturing thematic information, and a sentence-level latent variable to improve response diversity.

When modeling topic information, (Cui et al. 2022) introduced two distinct latent variables at the topic level to ensure dialogue consistency and role preference, as well as to enhance context sensitivity. The traditional approach to conversational chatbot development relies on retrieval-based systems, which require extensive databases to extract the most relevant parameters. In contrast, training a chatbot using deep learning techniques demands fewer resources and allows for the generation of dialog expressions. Hsueh and Chou proposed a method to develop a task-oriented chatbot using a sentence generation model based on an adversarial generative network. The model architecture consists of a generator, which creates diverse sentences, and a discriminator, which evaluates the sentences by comparing the generated ones to the ground truth. Within the generator, the attention mechanism is integrated with a sequence model, utilizing hierarchical short-term memory to improve sentence generation.

The discriminator mechanism assigns low rewards to repeated phrases and high rewards to diverse phrases. This system has enabled the generation of more varied and informative sentences (Hsueh and Chou 2023). Cheng et al. proposed a Convolutional Neural Network framework to extract and evaluate suspicious skin regions. The framework involves the following phases: (i) Image collection and resizing, (ii) Suspicious skin section extraction using VGG-UNet, (iii) Deep-feature extraction, (iv) Mining handcrafted features from the suspicious skin section, (v) Serial feature integration, and (vi) Classifier training and validation (Cheng et al. 2022). Hongming and Xiaochun proposed a cognitive diagnosis model based on multi-level attribute scores, designed to assess students' resource mastery levels based on their answers and the relevance of knowledge points (Guo and Cheng 2018).

Wang et al. introduced a PSSPNN model for classifying COVID-19, secondary pulmonary tuberculosis, community-captured pneumonia, and healthy subjects. The PSSPNN includes five improvements: (1) The introduction of the n-conv stochastic pooling module, (2) The development of a stochastic pooling neural network, (3) The incorporation of PatchShuffle as a regularization term, (4) The application of multiple-way data augmentation, and (5) The use of Grad-CAM to interpret the AI model (Wang et al. 2021).

3. GENERAL ARCHITECTURE OF THE PROPOSED CHATBOT

We have proposed a hybrid model that combines rule-based reasoning and machine learning. Rule-based chatbots use rules to create responses to answer patient questions. The number of answers is closely linked to the number of rules (Berghout et al. 2020). By using the rules, we can control the responses. Knowing that a doctor expert in this disease creates these answers, we can achieve better and more effective results. Several chatbots use this approach; we cite, for example, The ELIZA program, which was designed to simulate the discussion of a psychiatrist with patients (Sharma et al. 2022). Another example is the PARRY chatbot which was developed to simulate a paranoid patient. The rules-based approach is straightforward to implement but suffers from several problems, such as the lack of understanding of user questions (Shum, He, and Li 2018). Our proposal is also based on machine learning. We call chatbots following this approach "Self-Learning-Based.". The main advantage is that rules do not limit the number of questions. We distinguish two subcategories of this approach. The first is a retrieval-based chatbot that ranks data by a reward system. This type is the most used since it achieves the objectives in most applications. The second one is Generation-based. The principle of this type is to generate responses instead of selecting one from a database. So, this type allows proposing an answer to any question. But the answer quality is usually nonsense and does not respect grammar (Xu et al. 2021).

The chatbot architecture is linked to the domain of use. The architecture design is a long step because it requires answering several questions. For example, does the application allow the use of asynchronous mode? Another fascinating question is to determine whether we can add new data to the learning base or not. The answer to these questions will allow us to know the most appropriate techniques for our chatbot.

The main part of this architecture is the environment. At this level, user queries are interpreted using advanced machine learning algorithms to identify the meaning of the user's query and match it to the list of available responses supported by the software. In this part, we also manage the context of distinction by generating a chronological order of expressions. An essential part of the architecture is the management of the dialog. It is based on the Feedback Mechanism, which controls time and user responses to improve learning by exploiting errors to correct the model. The topmost layer of this part is Policy Learning which allows the bot to learn using the history of other users.

At the level of the questions and answers subsystem, the chatbot interprets and proposes the answers to the users. We can use manual training based on the definition of the answers by the expert or automated training, which consists of submitting different documents, and the system proposes a list of questions and answers from these documents.

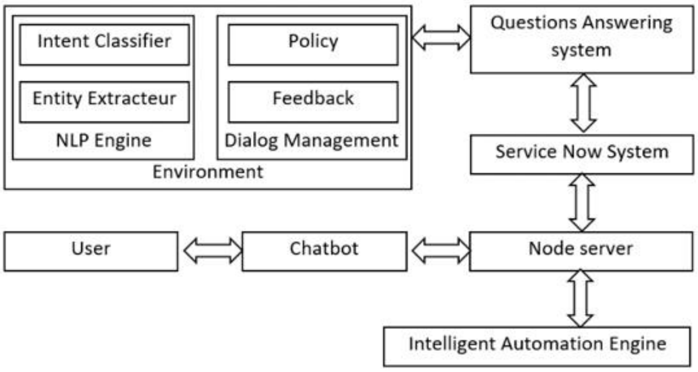

Figure 2. Different components of a chatbot regardless of the used field

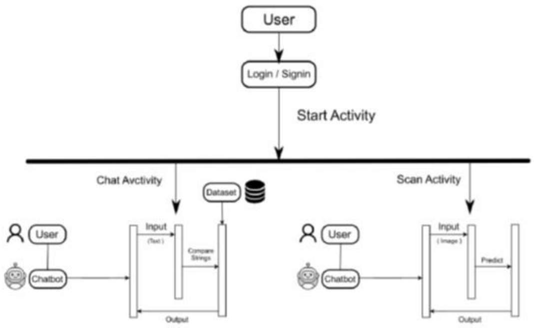

Figure 3. Activity diagram of the medical chatbot

Another important element is the set of APIs that provide several chatbot functionalities. A group of servers ensures communication with users. At the user level, various client applications can be considered a platform for chatbot interfaces.

4. AI TOOLS

It is very useful to create a textual medical chatbot to participate in the fight against the COVID-19 epidemic (Mehfooz et al. 2021). Our proposal tries to provide answers to frequently asked questions. It is also able to adapt to any new question concerning the coronavirus. By creating this software, we have achieved two objectives ensuring a good diagnosis and reducing the infection rate. Our proposal's two technological pillars are natural language processing and machine learning.

4.1 Neural Networks

Neural networks are used in several fields to achieve various tasks. The majority of pattern recognition applications are based on ANN. Significant progress is being made in natural language processing thanks to ANNs (Mellia et al. 2021). The most popular operating principle is supervised machine learning, which is based on the use of several layers. This architecture makes it possible to determine the complex relationships between the different variables of the system. This technique was inspired by the natural reasoning system of human beings. ANN is a robust model that provides classification and regression (Mutiwokuziva et al. 2017). The essential elements of this technique are :

- Inputs that are used to predict outputs.
- Weights are used to represent the importance of a value in the prediction process.
- Bias is used to move the activation function left or right.
- An activation function is used to give non-linearity in the process.

The learning process uses a loss function that compares the found value to the desired value. This operation is repeated several times to minimize the discrepancy between these values. At the end of each iteration, the networks' weights are set. Several algorithms provide this function, such as gradient descent (Mittal et al. 2021).

4.2 Convolutional neural networks

Like any other neural network, CNNs consist of several layers of neurons that ensure hierarchical learning (Fig. 4). With technological evolution and the appearance of specialized processing units such as Tensor Processing Units, deep learning has experienced tremendous progress and is becoming the most widely used model for classification and regression. In addition to the training layers of an ordinary ANN model, CNN offers a set of layers that ensures the extraction of the characteristics specific to an observation by applying a compression function. With the principles of this step and the application of several filters, the outputs are merged into a single data structure.

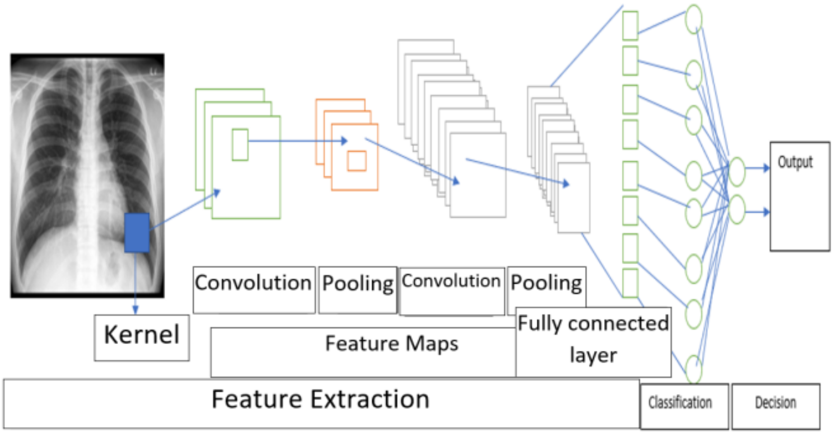

Figure 4. General architecture of the CNN classifier

In the convolution operation, we started by defining the filter window size. This window moves progressively from one side to another by several steps until the entire data structure is traversed. For each piece of data, a feature map is calculated, making it possible to locate the relevant features in the observation. Another significant step in CNN is the Max-Pooling, which reduces the data structure size resulting from the convolutional step. By applying this step, the learning time is reduced. Knowing its principle is to calculate the maximum of each structure region. The result is a data structure with the same form as the inputs (Ni 2021).

4.3 Plugins and APIs.

The four essential concepts used during the development steps are Activities, Content Providers, Intents, and Services (Hasan et al. 2021). We used Chaquopy, a Python plugin for Android's Gradle-based build system, to develop our system. To store chat messages instantly, we used Firebase. It is a powerful tool developed by Google for this purpose.

In the natural text processing part, NLTK (Natural Language Toolkit) proved to be extremely helpful as it provided a simple interface for using lexical resources and text analysis libraries. It also supports classification, tokenization, lemmatization, tagging, parsing, and semantic reasoning. To implement our intelligent model, we utilized TFlearn, a library built on top of TensorFlow. TensorFlow, created by the Google Brain team, is an open-source library for machine learning and deep learning (Tang 2016). The transfer of objects is carried out using the same methods as in JAVA: serialization and deserialization. Pickle handles both processes by converting Python objects into a stream of bytes for local storage or network transfer (Fasnacht 2018).

5. IMAGES CLASSIFICATION

5.1 Data pre-processing

Researchers from Qatar, Bangladesh, Pakistan, and Malaysia have compiled a comprehensive chest X-ray image database, which includes 3,616 images of COVID-19-positive cases, 10,192 images of normal cases, 6,012 images of lung opacity cases, and 1,345 images of viral pneumonia cases. To preprocess the images, each image is normalized by dividing it by the standard deviation of the dataset. This standardization process helps to ensure that the images have a consistent scale, improving the performance and accuracy of the machine learning models trained on this dataset. This operation aims to make the mean of data between 0 and std. Then we rotate all images by a 10° angle range. Our model assigns the value 0.2 to random zoom and brightness offset. Finally, we split the data into two subsets 70% training and 30% validation.

5.2 Classifier configuration

Thus, our model contains four blocks in total, the flattened, fully connected layer and the classifier, which includes the densest and the SoftMax activation function.

- First block: convolutional layer, LeakyReLU, max-pooling.
- Second block: Convolutional layer, LeakyReLU, max-pooling.
- Third block: Convolutional layer, LeakyReLU.
- Fourth block: convolutional layer, LeakyReLU, max-pooling.

The Conv2D layer creates a convolution kernel that processes the layer's input to generate tensor outputs. When used as the first layer in a model, its configuration will differ slightly from other layers. Specifically, the number of output filters in the convolution must be specified. In this case, we used eight filters for blocks 1 and 2, and 16 filters for blocks 3 and 4. For the height and width of the 2D convolution window, the first block uses a kernel size of (5,5), while the remaining blocks use a kernel size of (3,3). Padding was not applied to this layer across the four blocks. For activation, we used Leaky ReLU (Leaky Rectified Linear Unit), which is a variation of the standard ReLU function. Leaky ReLU introduces a small, non-zero slope for negative values instead of a flat zero, preventing the "dying ReLU" problem. This helps maintain a slight gradient even when the unit is inactive, which is beneficial for training by mitigating the effects of sparse gradients:

$$f(x) = \begin{cases} \propto\times x & if\ x < 0 \\ x & if\ x \geq 0 \end{cases} \quad (1)$$

In our model, $\propto = 0.1$ for all blocks. We used the MaxPooling2D function, which takes the maximum value over an input window for each channel. The window is shifted by a step of (2,2) along each dimension. Table 1 displays the sequential model layers and the output shape of each layer.

Table 1. Proposed sequential model

Layer (type)	Output Shape	Parameters
Conv2d	(None, 256, 256, 8)	608
Leaky_re_lu	(None, 256, 256, 8)	0
Maxpooling2d	(None, 256, 256, 8)	0
Conv2d_1	(None, 128, 128, 8)	584
Leaky_re_lu_1	(None, 128, 128, 8)	0
Maxpooling2d_1	(None, 64, 64, 8)	0
Conv2d_2	(None, 64, 64, 16)	1168
Leaky_re_lu_2	(None, 64, 64, 16)	0
Conv2d_3	(None, 64, 64, 16)	2320
Leaky_re_lu_3	(None, 64, 64, 16)	0

We used the Adam optimizer, which is a stochastic gradient descent method that adapts the learning rate by estimating the first and second-order moments of the gradients. The Adam optimizer combines the advantages of two other extensions of stochastic gradient descent, namely AdaGrad and RMSProp, to compute adaptive learning rates for each parameter. In our case, the first and second-order moments were fixed at 0.9. We set the learning rate to 0.001 to optimize the training process efficiently and avoid issues such as overshooting during gradient updates.

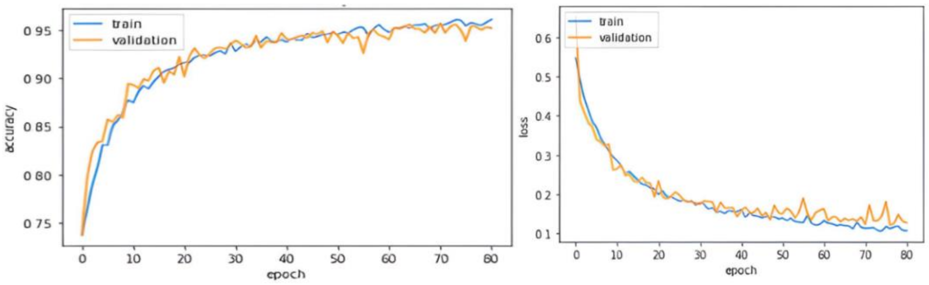

Figure 5. Model accuracy and loss

The time complexity of the proposed algorithms is O(Im), where I is the iteration number, and m is the dimension of text or image. We can reduce the complexity by using a features-selection procedure. In the worst case, all the features are selected. It will result in m evaluations. After "I" iterations, the classification procedure will be evaluated "Im" times.

5.3 Training the model

We divided the pre-processed dataset into groups of 32 items each. Then, we initialized the different environment variables. This model will train in 100 epochs, so our model needs to fit the training data in 100 cycles through the entire training dataset. Early termination stops training when the accuracy (monitored metric), as shown in fig.6, has stopped improving. Thus, we stop training when the model starts to overfit (accuracy drops). We used another stopping criterion, patience, which is the number of epochs without improvement.

5.4 Fitting the model

First, we created two functions to perform training and validation. The stopping criterion is defined in the function code. Another function was created to ensure early shutdown. The use of multithreading is An exciting property while training to access this step. After the end of the learning phase, we notice that the learning stopped at epoch 81. It means that the model started to overfit after epoch 81. We notice that the accuracy reached 96.08%, which is an excellent score, and the loss is 10.79%. The next step is to save the model so that we can evaluate it later and reuse it for our Android application. First, we loaded the model from the saved file. Second, we run the prediction function. Third, importing and using six performance metrics (confusion matrix, accuracy score, precision, recall, f1 score, and ROC AUC score) from the SciKit-learn library.

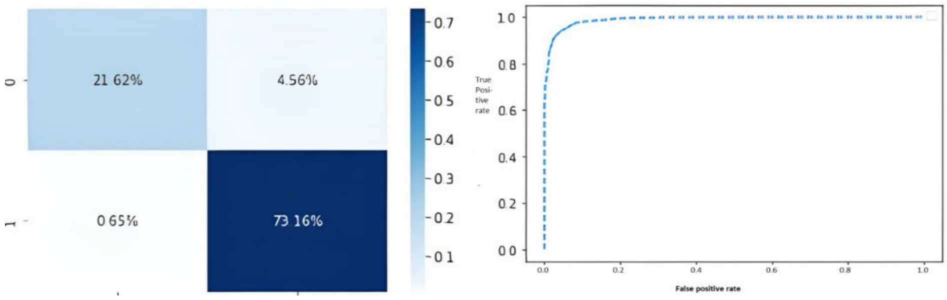

Figure 6. Confusion matrix using seaborn and Evaluation accuracy

Finally, we drew the ROC curve using Matplotlib and the confusion matrix using seaborn in fig.7. We notice that the evaluation accuracy shown in fig.8 is lower than the training accuracy. And now, we can test our model in real application after completing all the above steps.

6. THE CHATBOT MODEL

6.1 Presentation

In our code, the data is a JSON file of intents and groups of possible questions and answers. Data pre-processing includes the steps needed to transform or encode the data question so that the machine can analyze it smoothly. In order for a model to be accurate and precise in predictions, the algorithm must be able to easily interpret the characteristics of the data. We need to remove all unnecessary features and repeated words in the code using the NLTK "word_tokenize" method. The next step is to remove the morphological affixes from the words, leaving only the root of the word and all lowercase characters.

Because we have a string of data and the neural network only uses numbers, we need to create a bag of words containing the frequency of our input. After defining the form of inputs used by our model, we create the layers. Next, we define "SoftMax" as the activation function, and finally, we train our model and save it. The user types in a phrase and then creates a set of words to predict the answer and return the result. We notice that the model training phase is over, and the text classification model started to overfit after epoch 1000. We also noticed that the accuracy reached 99.77%, which is a perfect score. And the loss is 8.45%. The next step is to save the model so that we can evaluate it later and reuse it for our Android application.

6.2 Testing the models

To implement our system, we used models without metadata. We allocated a buffer of bytes, then acquired the user input as a list of arrays, and then passed the buffer of bytes to the input processing function. So basically, image classification is the same as text classification, but instead of a one-dimensional array, a list of arrays is used here. Among the input parameters, there is the size of the image. The predicted class represents the processing result.

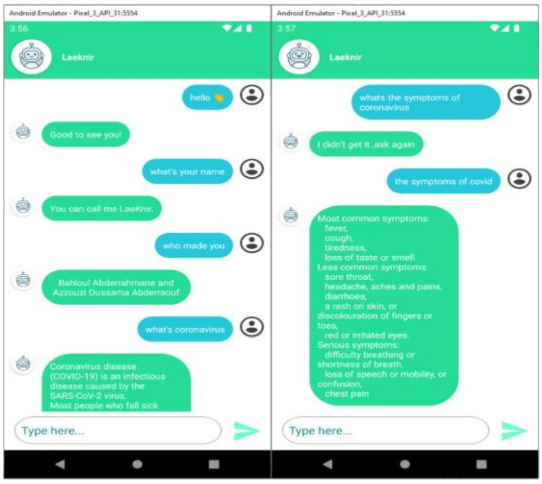

Figure 7. Chatbot (covid-19) activity

The home screen activity of our software contains the various functions offered by our application and some information about the user. Thus, the software requires an internet connection as all data is saved in a Firebase database. Our software provides two methods to download the X-ray image and gives the predicted class instantly. Here, the user needs to provide the necessary permissions to the app so that the application can access the images. Chatbot (covid-19) activity contains the second pattern(fig.11) and expects a text or string; when the user writes a message or question to the bot, the process will take less than a second to find and display the appropriate response. We have not considered all possible questions. If the system does not find the class of the question, it asks the user to reformulate his question.

7. DRUGS CONSUMPTION

For several centuries, leaders have discovered the importance of information to have an idea of their resources. In the 17th century, the calculation of probabilities and statistical methods appeared in Europe. Several researchers have participated in this field, we cite for example Pascal, Bernouli, Moivre, Laplace, Gauss, Mendel, Pearson, Fisher, etc…

Currently, we can easily notice the use of statistical methods applied in different fields such as medicine, agronomy, sociology, and industry.

The statistical study is based on the observation of variation, which consists of gathering coded data. Statistical methods are also based on the collection of data either exhaustively or by sampling. This data is sorted and organized into tables or diagrams. The results are interpreted by comparing them with those deduced from probability theory. To be able to make a decision, we must evaluate a statistical quantity such as the mean or the variance (estimators, confidence intervals). We also need to know if the two populations are comparable (hypothesis testing). Finally, we determine if two quantities are related and in what way (correlation, analytical adjustment).

Time series are widely used across various domains, and in this study, we focus on their application in the medical field (Merahi and Bibi 2021). The primary aim is to explain the evolution of a phenomenon over time. Another key objective is to predict the future behavior of this phenomenon. In this context, we have a series of observations taken at different time points, which consists of numerical values indexed by time. These observations provide valuable insights into trends, patterns, and future projections, making time series analysis an essential tool in medical research and practice. The set of observations of a group of real random variables denoted $(X_t)_{t \in \Theta}$ or X_t, $t \in \Theta$ represents a chronological (or temporal) series, where the set Θ represents a space of time which can be continuous or discreet. Schematically, the main steps in processing a time series are as follows: data correction, observation of the series, modeling, analysis of the series from its components, model diagnosis, and prediction.

The pre-processing phase in the study of a time series consists of processing and modifying the raw data. We cite for example the treatment of missing values, the construction of sub-series, the standardization in order to determine the size of the intervals, and the transformation of data for various reasons for example in economics, we use Box-Cox transformations.

A general rule in descriptive statistics is to start by observing the data before moving on to the time series modeling phase. For the same phenomenon, several models can be proposed. There are mainly two types of models. Deterministic models based on descriptive statistics. In this category, the most used models are the additive model and the multiplicative model and intervene for example in the ARCH and GARCH models (Atkinson, Riani, and Corbellini 2021). At the end of the modeling, the characteristics of the model are treated in a meticulous way. The verification phase comes after the construction of the model and the estimation of the parameters. Verification consists of analyzing residuals and conducting tests. Before diving deeper into the study, it is crucial to examine all observations in a time series. The seasonal variation S is simply added to the trend Z in the additive model, while in the multiplicative model, the seasonal variation S is proportional to the trend Z. This distinction can be made using either a graphical or analytical method. After completing these steps, predictions can be made based on the patterns and trends identified in the time series data.

8. DATA STATISTICAL ANALYSIS

At the beginning of this section, we will analyze the total quantities of drugs consumed. For this, we will explain the relationship between the number of drugs and the number of patients. We note that the number of patients registered in the entrance department of the Mohamed Boudiaf Hospital increased during the years (2018 by 27%) and (2019 by 31%) then this number decreased during the years (2020 by 21%) and (2021 by 21%). This decrease during the year 2020 is caused by the outbreak of the COVID-19 pandemic. On February 25, 2020, the Algerian Minister of Health, Abderrahmane Benbouzid announced to the press the appearance of the first case of Covid-19. This is a foreign employee from Lombardy (an Italian city). This city represents one of the most affected areas in Italy. Due to the outbreak of the COVID-19 pandemic, the hospital stopped operations surgeries except for emergency cases, and parents have stopped bringing their children to the hospital unless necessary and serious for fear of infection with the COVID-19 virus. In general, all citizens avoid going to the hospital. Although the number of patients has reduced, the number of anticoagulants and recommended drugs and analgesics, and anti-inflammatories increased; because COVID-19 disease requires treatment with this drug.

The amount of anticoagulant and antibiotic and analgesic and anti-inflammatory drugs has increased because the disease COVID-19 requires treatment with these drugs. Anticoagulants (2018/15%), (2019/16%), (2020/26%), (2021/43%); Anti-inflammatories and analgesics (2018/11%), (2019/23%), (2020/30%), (2021/36%); Antibiotics (2018/14%), (2019/14%), (2020/33%), (2021/39%). The proportionality between the number of patients and the quantity of drugs consumed has also increased. Anticoagulants 2018 (15/27=0.55), 2019 (16/31=0.52), 2020 (26/21=1.24), 2021 (43/21=2.04), Anti-inflammatories and Analgesics 2018 (11/27=0 .41), 2019 (23/31=0.74), 2020 (30/21=1.43), 2021 (36/21=1.71); Antibiotics 2018 (14/27=0.52), 2019 (14/31=0.45), 2020 (33/21=1.57), 2021 (39/21=1.86).

These are drugs that inhibit the natural coagulability of the blood but will keep the same viscosity. This type of medication prevents clots from building up in the blood vessels and prevents the onset of dangerous diseases such as pulmonary embolism and cerebral embolism. This anticoagulant treatment is recommended for temporarily immobilized patients. Antibiotics are natural or synthetic drugs. They act on bacteria or protozoa. they can kill them or prevent their proliferation. They have no effect on viruses.

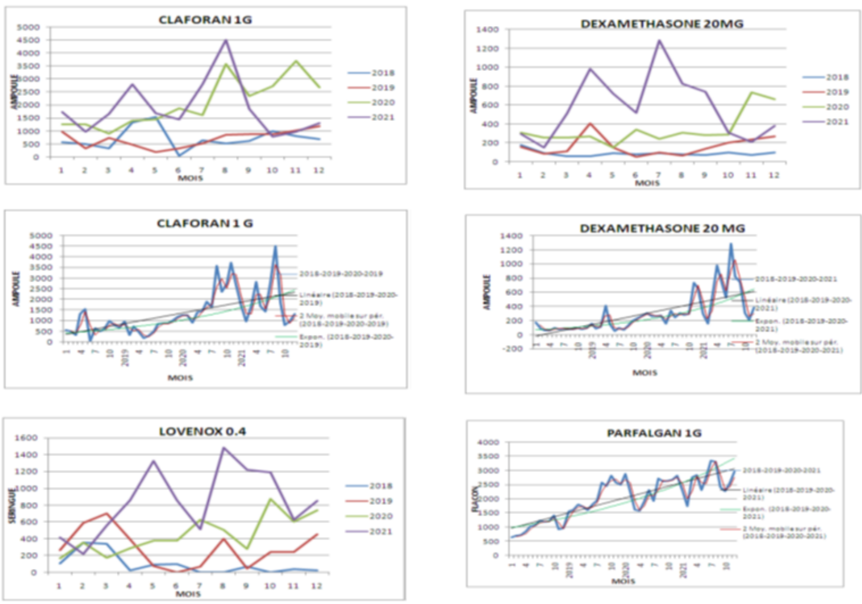

Figure 8. Example analysis of analgesics, Anti-inflammatories, Antibiotics, and Anticoagulants

Anti-inflammatories are medications that help reduce inflammation, regardless of the underlying cause. While they do not target the root cause of the inflammation, they alleviate its symptoms, including pain. There are two main categories of anti-inflammatory drugs: nonsteroidal anti-inflammatory drugs (NSAIDs) and steroidal anti-inflammatory drugs, commonly known as corticosteroids.

Analgesics are drugs used to relieve pain. They act directly on the nervous system to interrupt pain signals. Analgesics do not act on the cause of the pain. Analgesics are classified into three levels, from the least powerful to the most powerful. This makes it possible to choose the drug according to the intensity of the pain.

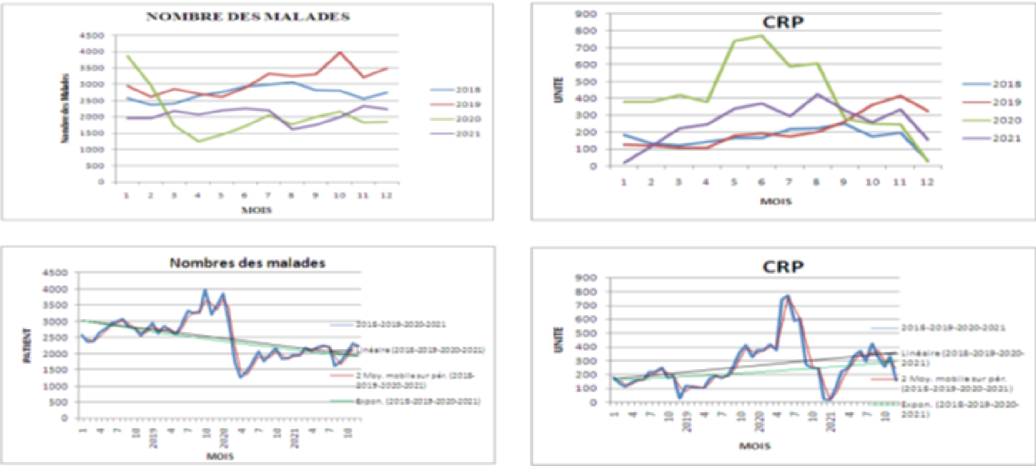

Figure 9. Graphs of biological assessments

Based on the number of patients received in the different departments of the hospital, we note that during the Delta wave, the Covid department received the largest number of patients. Figure 10 presents the biological assessments carried out at the hospital level. The purpose of a check-up is to confirm a diagnosis and improve treatment and monitor a patient's state of health.

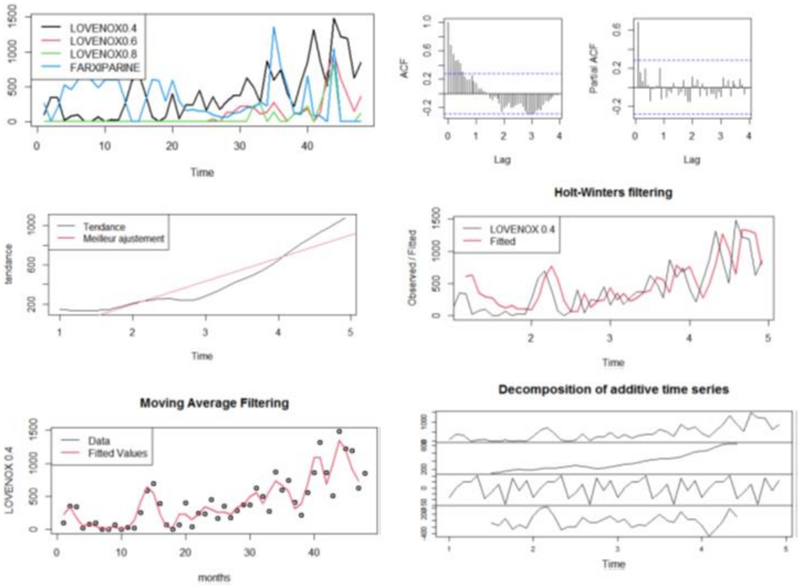

Figure 10. Consumption prediction of LOVENOX 0.4

According to Figure 9, we observe that during the years 2018 and 2019 the curves for the number of patients and the consumption of drugs are almost identical. but during the years 2020 and 2021, the curves are opposite, the quantities of drugs consumed increased and the number of patients decreases due to Covid-19. We note that the quantities of drugs appear in the form of peaks during the Covid-19 waves.

The most violent wave of Covid-19 took place in August 2021, and due to lack of oxygen, Mohamed Boudiaf Barika Hospital lost a large number of Covid-19 patients, especially on an unforgettable day on 04/08 /2021 the death toll reached around 20. it is observed that all the graphs of the quantities of the drugs form an increasing trend curve. that is to say, the consumption increased from the year 2018 until 2021, in parallel, the graph of the numbers of the sick forms descending trend curves. We used the moving average (MA) technique to calculate the overall trend in the data set. This technique is very useful for predicting short-term trends. The forecast curve and results are given as follows:

We applied another method called Double Exponential Smoothing. It belongs to the type of exponential forecasting methods. It allows for weighted averages where more weight can be given to recent observations and less weight to older observations. Exponential smoothing methods are intuitive, computationally efficient, and generally applicable to a wide range of time series. The forecasts at the horizon h = 12 months are: (838.4155, 838.8846, 839.3538, 839.8229, 840.2920, 840.7612, ,841.2303, 841.6994, 842.1686, 842.6377, 843.1068, 843.5760). Here is the analysis of the trend by a linear model:

- Residuals: Min 1Q Median 3Q Max; -156,261 -92,114 3,804 79,061 187,903
- Coefficients: Estimate Std. Error t value Pr(>|t|)
- (Intercept) -273.6 41.6 -6.577 3.94e-08 ***
- time(trend) 235.8 13.1 18.000 < 2e-16 ***
- Meaning. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
- Residual standard error: 104.8 on 46 degrees of freedom
- Multiple R-squared: 0.8757, Adjusted R-squared: 0.873
- F-statistic: 324 on 1 and 46 DF, p-value: < 2.2e-16

We applied the Box-Jenkins method based on the AR(p), MA(q), ARMA(p,q), ARIMA(p,d,q) and SARIMA(p,d,q) models to the prediction problem of drug consumption. We find that the simple correlogram of the series decreases linearly towards 0. We can say that there is the persistence of a trend which implies that the series is not stationary. We applied the Box-Pierce test. Knowing that the null hypothesis of this test is the non-stationarity of the studied series. The result of the test gives us:

X-squared=22.332, df=1, p-value=2.293e-06. The best way to make it stationary is to apply the difference operator $\Delta = (1 - B)$ of order d. After having differentiated our series using the difference operator for d = 1, we apply the Box-Pierce test on the new series obtained to check if it is stationary. X-squared=3.4562, df=1, p-value=0.06301.

We find that the order of derivation, of the ARIMA and SARIMA models, d is equal to one. The estimation of p and q is done by observing the correlogram of the autocorrelation and partial autocorrelation functions. Knowing that d = 1 is the order of the differential function. The simple correlogram of the autocorrelation function (ACF) gives us the value of q. The partial correlogram of the partial autocorrelation function gives the value of p.

We note that on the simple correlogram ACF indicates that the significant term Q is equal to two. The PACF partial correlogram indicates that the significant term P is equal to two.

To choose the final model among the models studied, we used the AIC criterion. So, the best model is the one that minimizes the AIC criterion. In this case, the best models are:

- SARIMA(1, 1, 2): AIC=651.677; SARIMA(2, 1, 2): AIC=653.0625,
- ARMA(1, 1): AIC=658.5522 ; SARIMA(1, 1, 1): AIC= 659.2479,

In this case, the best SARIMA(1, 1, 2), the results of the parameter estimation are reported in the following lines: arima(x = y1, order = c(1, 1, 2), seasonal = list(order = c(1, 1, 2), period = 12)).

- Coefficients: ar1 ma1 ma2 sar1 sma1 sma2
- 0.0122 -0.4354 -0.4186 0.2653 -0.5442 0.1437
- w.s. NaN 0.0773 0.1499 NaN 0.0830 NaN
- sigma2 estimated as 84664: log likelihood = -249.4, aic = 512.79.

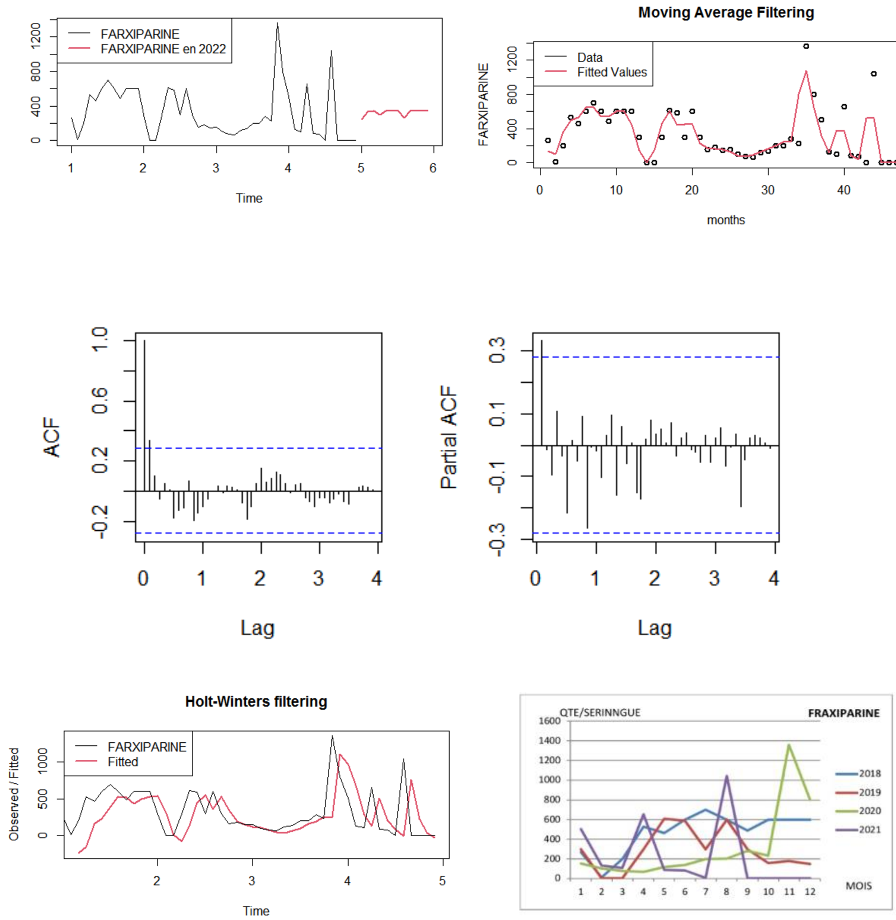

Figure 11. Consumption Prediction of FARXIPARINE

The expected values in 2022 of the consumption of LOVENOX 0.4 are reported in table 2.

Table 2. The consumption of LOVENOX 0.4 in 2022.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Consumption	711	661	862	1105	1449	1115	934	1614	1350	1499	1015	1226

To validate the SARIMA(1,1,2) model, we applied the Box-test. The null hypothesis on the residuals is white noise which gives a p-value =0.2613>alpha=0.01. Here is the analysis of the trend by a linear model: Call: lm(formula = trend time(trend))

- Residuals: Min 1Q Median 3Q Max
- -88.88 -52.29 -19.39 50.85 147.76
- Coefficients: Estimate Std. Error t value Pr(>|t|)
- (Intercept) -200.482 27.267 -7.353 2.70e-09 ***
- time(trend) 106.896 8.586 12.450 2.47e-16 ***
- Meaning. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
- Residual standard error: 68.68 on 46 degrees of freedom
- Multiple R-squared: 0.7711, Adjusted R-squared: 0.7662
- F-statistic: 155 on 1 and 46 DF, p-value: 2.468e-16.

By applying the Box-Jenkins method, we found the following results: d = 1, P = 1, Q = 2. By applying the AIC criterion, the best model is SARIMA(1, 1, 2). Using this model, the predicted values in 2022 of the consumption of "LOVENOX 0.6" are:

612.5578, 721.3200, 492.2708, 358.9583, 221.2253, 519.8340, 1128.1013, 1320.1497, 884.8963, 626.7022, 478.5788, 718.2017.

9. CONCLUSIONS

In this article, we propose a medical diagnostic system composed of two intelligent components based on TensorFlow Lite. Focusing on the prevalent and urgent issue of COVID-19, we aim to provide technical solutions to aid individuals during the pandemic. Through an intelligent agent, we offer guidance to help navigate these challenging circumstances. The second part of our study delves into combating COVID-19 through data analysis.

Our primary objective is to understand the relationship between various disease parameters and the consumption rates of specific drugs. Initially, we analyze pandemic-related information and subsequently develop models to identify factors influencing its spread. The insights gained from our findings will contribute to enhancing existing strategies. The data for the second part of our study was collected at Mohamed Boudiaf Hospital in the city of BARIKA, Algeria, covering patient numbers and drug consumption from 2018 to 2021. We observed a decline in drug consumption as the COVID-19 pandemic receded, returning almost to normal levels. Conversely, during peak periods of the four waves of COVID-19, drug consumption increased significantly. Future studies will involve exploring larger datasets to enhance accuracy.

Our diagnostic model can potentially be adapted for other diseases, and additional language support will broaden its accessibility. Integrating graphical tools will facilitate better communication with patients. Furthermore, we will introduce a neural network optimizer utilizing ant colony optimization to minimize inputs while increasing accuracy. Training using TPU will reduce computation time, and the implementation of deep reinforcement learning will expand the training database, leading to more precise results.

ACKNOWLEDGMENTS

This work was supported by a grant from the Ministry of Higher Education and Scientific Research, PRFU, project no. C00L07UN050220200003.

REFERENCES

[1] Amer, Eslam, Ahmed Hazem, Omar Farouk, Albert Louca, Youssef Mohamed, and Michel Ashraf. (2021). “A Proposed Chatbot Framework for COVID-19.” In 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), 263–68. IEEE. https://doi.org/10.1109/MIUCC52538.2021.9447652.

[2] Anki, Prasnurzaki, Alhadi Bustamam, and Rinaldi Anwar Buyung. (2021). “Comparative Analysis of Performance between Multimodal Implementation of Chatbot Based on News Classification Data Using Categories.” Electronics 10 (21): 2696. https://doi.org/10.3390/electronics10212696.

[3] Atkinson, Anthony C., Marco Riani, and Aldo Corbellini. (2021). “The Box–Cox Transformation: Review and Extensions.” Statistical Science 36 (2). https://doi.org/10.1214/20-STS778.

[4] Berghout, Tarek, Leïla-Hayet Mouss, Ouahab Kadri, Lotfi Saïdi, and Mohamed Benbouzid. (2020). “Aircraft Engines Remaining Useful Life Prediction with an Adaptive Denoising Online Sequential Extreme Learning Machine.” Engineering Applications of Artificial Intelligence 96 (November): 103936. https://doi.org/10.1016/j.engappai.2020.103936.

[5] Chebout, Mohamed Sedik, and Oussama Kabour. (2022). “Machine Learning-Based Forecasting Models for COVID-19 Spread in Algeria.” Model Assisted Statistics and Applications 17 (2): 99–108. https://doi.org/10.3233/MAS-220013.

[6] Cheng, Xiaochun, Seifedine Kadry, Maytham N. Meqdad, and Rubén González Crespo. (2022). “CNN Supported Framework for Automatic Extraction and Evaluation of Dermoscopy Images.” The Journal of Supercomputing 78 (15): 17114–31. https://doi.org/10.1007/s11227-022-04561-w.

[7] Cui, Fuwei, Hui Di, Hui Huang, Kazushige Ouchi, Ze Liu, and Jinan Xu. (2022). “Hierarchical Latent Variables Structure for Topic Aware Multi-Turn Conversation.” Journal of Intelligent & Fuzzy Systems 43 (3): 3805–14. https://doi.org/10.3233/JIFS-211886.

[8] Dhyani, Manyu, and Rajiv Kumar. (2021). “An Intelligent Chatbot Using Deep Learning with Bidirectional RNN and Attention Model.” Materials Today: Proceedings 34: 817–24. https://doi.org/10.1016/j.matpr.2020.05.450.

[9] Fan, Xiangmin, Daren Chao, Zhan Zhang, Dakuo Wang, Xiaohua Li, and Feng Tian. (2021). “Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study.” Journal of Medical Internet Research 23 (1): e19928. https://doi.org/10.2196/19928.

[10] Fasnacht, Laurent. (2018). “Mmappickle: Python 3 Module to Store Memory-Mapped Numpy Array in Pickle Format.” Journal of Open Source Software 3 (26): 651. https://doi.org/10.21105/joss.00651.

[11] Guo, Hongming, and Xiaochun Cheng. (2018). “Individual Recommendation Method of College Physical Education Resources Based on Cognitive Diagnosis Model.” ICST Transactions on Scalable Information Systems, July, 173379. https://doi.org/10.4108/eai.10-2-2022.173379.

[12] Hasan, Rajibul, Bernadett Koles, Mustafeed Zaman, and Justin Paul. (2021). “The Potential of Chatbots in Travel and Tourism Services in the Context of Social Distancing.” International Journal of Technology Intelligence and Planning 13 (1): 63. https://doi.org/10.1504/IJTIP.2021.117998.

[13] Hsueh, Yu-Ling, and Tai-Liang Chou. (2023). “A Task-Oriented Chatbot Based on LSTM and Reinforcement Learning.” ACM Transactions on Asian and Low-Resource Language Information Processing 22 (1): 1–27. https://doi.org/10.1145/3529649.

[14] Mehfooz, Fahad, Sakshi Jha, Sahil Singh, Shreya Saini, and Nidhi Sharma. (2021). “Medical Chatbot for Novel COVID-19.” In , 423–30. https://doi.org/10.1007/978-981-15-8354-4_42.

[15] Mellia, Joseph A., Marten N. Basta, Yoshiko Toyoda, Sammy Othman, Omar Elfanagely, Martin P. Morris, Luke Torre-Healy, Lyle H. Ungar, and John P. Fischer. (2021). “Natural Language Processing in Surgery.” Annals of Surgery 273 (5): 900–908. https://doi.org/10.1097/SLA.0000000000004419.

[16] Merahi, Fateh, and Abdelouahab Bibi. (2021). “Evolutionary Transfer Functions Solution for Continuous–Time Bilinear Stochastic Processes with Time-Varying Coefficients.” Communications in Statistics - Theory and Methods 50 (22): 5189–5214. https://doi.org/10.1080/03610926.2020.1726390.

[17] Mittal, Mamta, Gopi Battineni, Dharmendra Singh, Thakursingh Nagarwal, and Prabhakar Yadav. (2021). “Web-Based Chatbot for Frequently Asked Queries (FAQ) in Hospitals.” Journal of Taibah University Medical Sciences 16 (5): 740–46. https://doi.org/10.1016/j.jtumed.2021.06.002.

[18] Mutiwokuziva, Milla T, Melody W Chanda, Prudence Kadebu, Addlight Mukwazvure, and Tatenda T Gotora. (2017). “A Neural-Network Based Chat Bot.” In 2017 2nd International Conference on Communication and Electronics Systems (ICCES), 212–17. IEEE. https://doi.org/10.1109/CESYS.2017.8321268.

[19] Ni, Jiasheng. (2021). “A Medical Service Application Based on 3D-CNN and Knowledge Graph.” Journal of Physics: Conference Series 2078 (1): 012048. https://doi.org/10.1088/1742-6596/2078/1/012048.

[20] Sharma, Vibhor, Shashi Bhushan, Anuj Kumar Singh, and Pramod Kumar. (2022). “A Novel Secure Vector Product for Protecting the Privacy of Data in Vertically Partitioned Dataset.” In , 285–95. https://doi.org/10.1007/978-981-19-2828-4_28.

[21] Shum, Heung-yeung, Xiao-dong He, and Di Li. (2018). “From Eliza to XiaoIce: Challenges and Opportunities with Social Chatbots.” Frontiers of Information Technology & Electronic Engineering 19 (1): 10–26. https://doi.org/10.1631/FITEE.1700826.

[22] Tang, Yuan. (2016). “TF.Learn: TensorFlow’s High-Level Module for Distributed Machine Learning,” December. http://arxiv.org/abs/1612.04251.

[23] Temniranrat, Pitchayagan, Kantip Kiratiratanapruk, Apichon Kitvimonrat, Wasin Sinthupinyo, and Sujin Patarapuwadol. (2021). “A System for Automatic Rice Disease Detection from Rice Paddy Images Serviced via a Chatbot.” Computers and Electronics in Agriculture 185 (June): 106156. https://doi.org/10.1016/j.compag.2021.106156.

[24] Tsuru, Satoko, Tetsuro Tamamoto, Akihiro Nakao, Kouichi Tanizaki, and Naohisa Yahagi. (2022). “Patient Data Sharing and Reduction of Overtime Work of Nurses by Innovation of Nursing Records Using Structured Clinical Knowledge.” In . https://doi.org/10.3233/SHTI220514.

[25] Urquiza-Yllescas, José Fidel, Sonia Mendoza, José Rodríguez, and Luis Martín Sánchez-Adame. (2022). “An Approach to the Classification of Educational Chatbots.” Journal of Intelligent and Fuzzy Systems 43 (4): 5095–5107. https://doi.org/10.3233/JIFS-213275.

[26] Wang, Shui-Hua, Yin Zhang, Xiaochun Cheng, Xin Zhang, and Yu-Dong Zhang. (2021). “PSSPNN: PatchShuffle Stochastic Pooling Neural Network for an Explainable Diagnosis of COVID-19 with Multiple-Way Data Augmentation.” Edited by Martti Juhola. Computational and Mathematical Methods in Medicine 2021 (March): 1–18. https://doi.org/10.1155/2021/6633755.

[27] Xu, Ruijian, Chongyang Tao, Jiazhan Feng, Wei Wu, Rui Yan, and Dongyan Zhao. (2021). “Response Ranking with Multi-Types of Deep Interactive Representations in Retrieval-Based Dialogues.” ACM Transactions on Information Systems 39 (4): 1–28. https://doi.org/10.1145/3462207.

[28] Zhang, Zhe, Yiyang Zhang, Xiang Li, Yurong Qian, and Tao Zhang. (2022). “BMCSA:Multi-Feature Spatial Convolution Semantic Matching Model Based on BERT.” Journal of Intelligent & Fuzzy Systems 43 (4): 4083–93. https://doi.org/10.3233/JIFS-212624.