International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 6, pp 115-122, June-2023

https://doi.org/10.47001/IRJIET/2023.706018

Healthcare Chat Bot for Identifying Diseases and Providing Referrals Using Machine Learning

[1]Pamudu Ratnayake, [2]Hansani Bandara, [3]Oshini Cooray, [4]Chamathka Ariyarathna, [5]Suriyaa Kumari, [6]Ravi Supunya

[1,2,3,4]Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka
[5,6]Lecturer, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka

***Abstract* - Healthcare chatbots have the potential to revolutionize healthcare by providing accessible, affordable, and personalized medical advice and support to individuals. In this research paper, we propose a healthcare chatbot system that uses machine learning algorithms to identify diseases and provide appropriate referrals to patients. The proposed chatbot is designed to interact with patients via natural language processing (NLP) and answer their questions related to their symptoms and other relevant factors. The system uses a deep learning-based approach to analyze patient data and provide accurate and personalized recommendations. The proposed chatbot system was evaluated using a dataset of medical records from patients with various diseases such as dengue, influenza, nail diseases. The evaluation results showed that the proposed system achieved high accuracy in identifying diseases and providing referrals. Furthermore, the system was able to provide personalized recommendations based on patients' unique symptoms and other relevant factors. The proposed healthcare chatbot system has the potential to improve healthcare delivery by providing quick and personalized medical advice and support to patients. The system's ability to identify diseases like nail diseases, acne diseases, covid-19, dengue, influenza and provide first aid using machine learning algorithms can help healthcare providers diagnose diseases at an early stage and provide timely and effective treatment to patients. Additionally, the proposed chatbot system can be easily integrated into existing healthcare systems, making it accessible to a broader population.**

Keywords: Image Processing, Transfer Learning, Random-forest.

I. INTRODUCTION

The healthcare industry faces many challenges, including the rising cost of healthcare, shortages of medical professionals, and the need for more efficient and effective healthcare delivery. One solution to these challenges is the use of healthcare chatbots, which can provide accessible and personalized medical advice and support to patients. Healthcare chatbots are AI-powered systems that can interact with patients, understand their symptoms, and recommend appropriate medical advice and treatment.

In this research paper, we propose a healthcare chatbot system that uses machine learning algorithms and image processing to identify diseases and provide appropriate referrals to patients. The chatbot is designed to interact with patients via natural language processing (NLP) and can answer their questions related to their symptoms, medical history, and other relevant factors. The proposed chatbot system leverages deep learning algorithms to analyze patient data and provide personalized recommendations.

Since this is a mobile application, the proposed healthcare chatbot system has several advantages. It can help alleviate the burden on healthcare professionals, enabling them to focus on critical tasks while chatbots handle routine tasks. Additionally, the chatbot system can provide 24/7 accessibility to medical advice, making it more convenient for patients to receive care. The system can also be integrated with electronic medical records, enabling healthcare providers to access patients' medical history and provide more accurate diagnoses and treatments.

The first objective of this research paper is, Identify the disease whether it is covid-19 or influenza or dengue or virus fever with accuracy percentage by accepting both Sinhala and English inputs and support vocal as well as text. The emergence and rapid spread of infectious diseases have become a global health challenge in recent times. Diseases such as dengue, influenza, COVID-19, and viral fever are highly contagious and can cause significant morbidity and mortality. Early identification and diagnosis of these diseases are crucial in providing appropriate treatment and reducing the risk of further transmission. Machine learning algorithms have shown promising results in identifying and diagnosing diseases based on symptoms. Random forest algorithm has been widely used in medical research due to its ability to handle large and complex datasets and produce accurate predictions. In this research, we aim to identify and diagnose four common infectious diseases, namely dengue, influenza, COVID-19, and viral fever, using the random forest algorithm. We utilized a dataset of symptoms and employed natural

ISSN (online): 2581-3048
Volume 7, Issue 6, pp 115-122, June-2023
https://doi.org/10.47001/IRJIET/2023.706018

language processing techniques to process inputs in vocal and text formats based on Sinhala and English languages.

The second objective of this research paper is nail disease identification using image processing. A sizable fraction of the global population is afflicted with frequent medical problems known as nail diseases. These problems might range from less serious diseases like psoriasis or melanoma to more serious illnesses like benign disorders like onychomycosis (fungal infection of the nail). For rapid treatment and management of nail illnesses, early and precise diagnosis is crucial. Delayed diagnosis might result in permanent damage to the nail or the surrounding tissue. The integration of image processing and machine learning techniques has developed in recent years as a viable strategy for nail disease diagnosis. These techniques can find patterns and characteristics that are symptomatic of nail problems by examining photographs of the nail and surrounding tissue. The accuracy of the diagnosis has been increased by using transfer learning, a method in which a model that has already been trained is utilized to address a related issue. In this paper, we provide a method for diagnosing nail diseases that makes use of transfer learning and image processing. Convolutional neural networks (CNNs) were employed by our system to categorize photos into several nail illness categories and extract characteristics from a collection of nail images. Accuracy, precision, recall, and F1 score were some of the assessment measures we used to assess the effectiveness of our system. Our studies' findings show that our system performs better than current cutting-edge techniques and can effectively identify nail disorders. Our work demonstrates that image processing and machine learning approaches may be effective diagnostic tools for nail illnesses and have the potential to increase the precision and effectiveness of medical diagnosis. The suggested approach for identifying nail disorders may be improved upon and incorporated into existing healthcare infrastructure, allowing for quick and precise diagnosis of nail diseases and better patient outcomes.

The third objective of this research paper is, Identify the skin type of the face and the type of acne to make recommendations. Skin diseases are relatively common and currently have several treatments available. Yet, sometimes correctly diagnosing these diseases can be difficult due to the rigid, difficult-to-discriminate symptoms they present with. Since their recent development, Deep Neural Networks has started to perform better than other algorithms in practically every area. Acne is a common skin disorder or condition that typically develops as a result of bacterial infection, clogged skin sebaceous glands, and clogged hair shafts. It can be identified by the presence of blackhead, milia, rosacea, scar, tineafasia [1]. Between the ages of 12 and 24 years, at least 85% of Sri Lankan experience acne. Because of the difficulty and shame that acne causes, sufferers frequently endure emotional and social instability. About 9.8% of females and 9% of males experience acne, which is a higher prevalence in females. A person's face may experience several different forms of acne lesions at once. For the right prescription of treatment and the advancement of healthcare, early and accurate detection of those particular illnesses is crucial [2]. Identification of acne kinds is now a crucial and heavily studied issue in the field of dermatology [3]. The physical observation and acne calculation technique of diagnosis is a time-consuming, labor-intensive approach that depends on the skill and experience of the practitioner [4]. Also, this strategy calls for a lengthy training period. A range of image processing and machine learning approaches are currently being employed to precisely identify the skin issues caused by acne on the human face in an effort to solve this problem [5]. In order to identify the different types of acne in the skin during cosmetic surgery, VISIA is a widely utilized commercial instrument. By looking at multispectral photos, it is accomplished. Unfortunately, it is highly pricey, and the dermatologist must manually draw the region of interest (ROI) during the inspection, which is also very unreliable.

The fourth objective of this research paper is, Identification and recommendation of best-suited First Aid Instructions based on the situation. In day-to-day life many health problems are arisen such as headache, wounds, bone fractures etc. Sometimes most of these problems can be solved with few first aid actions. Most of the people do not know how to react in these day-to-day medical problems. Therefore, sometimes they will become into critical conditions. Until the patient is carrying to the hospital, some necessary actions to be taken to keep the patient in an optimal health condition. In order to give necessary first aid instructions for medical problems in day to day life, we are proposing first aid functionality in our introducing app. There people can enter a relevant input explaining the medical problem and the app will give the exact first aid instruction to be followed by the patient. In Sri Lanka, yet any mobile or web application has been not invented to give first aid instructions to people with a higher accuracy level. Many researches that are available shows that chatbots that have been invented so far are only based on with a specific area such as mental health, heart disease etc. In our first aid chatbot we are covering every possible day-to-day health problem and first aid instructions have been given through machine learning algorithms by computationally analyzing the first aid instruction data set.

ML can be used to provide customized suggestions for the first aid training depending on the distinct symptoms. This could boost the accuracy and effectiveness of first aid care, potentially saving lives in terrible situations.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 6, pp 115-122, June-2023

https://doi.org/10.47001/IRJIET/2023.706018

The remainder of this research paper is organized as follows. Section II provides an overview of related work in healthcare chatbot systems according to our research objectives. Section III describes the methodologies that we used in this research and Section IV presents the evaluation of the proposed system, discussing the results and limitations. Finally, Section V concludes the paper and outlines future research directions.

II. RELATED WORK

2.1 Identify the disease whether it is Covid 19, Influenza, Dengue or Virus Fever

Machine learning algorithms have been widely used in medical research for disease identification and diagnosis. For instance, this research paper [6] utilized a decision tree algorithm to diagnose dengue fever based on clinical and laboratory parameters. Their results showed that the decision tree algorithm can accurately diagnose dengue fever with a high degree of accuracy. Similarly, in here [7] developed a deep learning-based approach to diagnose COVID-19 from chest X-ray images. Their findings demonstrated that the deep learning algorithm can effectively differentiate COVID-19 from other respiratory diseases with high accuracy. Other studies have employed natural language processing techniques for disease diagnosis. For example, [8] utilized a support vector machine algorithm to diagnose liver diseases based on patient symptoms and lab test results. They showed that the support vector machine algorithm can effectively identify liver diseases with high accuracy. Moreover, this research [9] used natural language processing techniques to extract features from patient health records and employed a deep learning algorithm to diagnose pneumonia. Their findings suggested that the deep learning algorithm can accurately diagnose pneumonia based on patient symptoms and health records.

Several studies have also explored the use of machine learning algorithms for the diagnosis of infectious diseases. For example, according to this research [10] utilized a decision tree algorithm to diagnose dengue fever based on clinical symptoms and laboratory parameters. Their results demonstrated that the decision tree algorithm can accurately diagnose dengue fever with high sensitivity and specificity. Additionally, this research [11] used machine learning algorithms to diagnose COVID-19 from chest X-ray images. Their findings showed that the algorithms can effectively differentiate COVID-19 from other respiratory diseases with high accuracy.

2.2 Nail Disease Identification

The use of machine learning and image processing algorithms to recognize and diagnose different diseases has gained popularity in recent years. Image processing methods have been utilized in dermatology to find skin conditions including melanoma and psoriasis. However, employing image processing methods to detect nail disorders has received less attention.

One of the few studies in this area is the work of [12], who developed a computer-aided diagnostic system for nail psoriasis. The system used image processing techniques to analyze nail images and classify them as either healthy or diseased. The authors achieved an accuracy of 88.6% using support vector machines (SVM) as the classification algorithm.

Another study by [13] used a deep learning approach to diagnose nail diseases. The authors used a dataset of 1300 nail images and trained a convolutional neural network (CNN) to classify the images into different categories of nail diseases. The authors achieved an accuracy of 89.2% using the CNN.

Transfer learning has become a viable strategy to boost the effectiveness of machine learning algorithms in recent years. Transfer learning entails optimizing previously trained models for a particular task from a starting point. Dermatology is one area where transfer learning has been successfully used in picture categorization tasks.

In the context of nail disease identification, transfer learning has been used [14] to develop a nail disease diagnosis system. The authors used a pre-trained CNN model and fine-tuned it on a dataset of 540 nail images. The authors achieved an accuracy of 92.5% using transfer learning.

By combining image processing methods and transfer learning, we expand previous research on nail disease identification in this publication. We employ data preprocessing and feature selection approaches to extract pertinent features from a collection of more than 3000 nail photos. Then, using transfer learning, we hone a pre-trained CNN model to classify the photos into several nail disease subcategories. Our findings show how useful transfer learning is for identifying nail disorders and how this strategy may be applied in clinical settings for nail disease early detection and diagnosis.

2.3 Identify the skin type of the face and the type of acne to make recommendations

"Deep learning for classification of melanoma and benign lesions in dermatological images" [15] by Esteva et al. (2017) - This paper presents a deep learning approach for classifying skin lesions as either benign or malignant. It uses a convolutional neural network (CNN) architecture and achieves high accuracy on a large dataset of dermatological images.

ISSN (online): 2581-3048
Volume 7, Issue 6, pp 115-122, June-2023
https://doi.org/10.47001/IRJIET/2023.706018

"Automatic acne grading system using deep convolutional neural networks" by Kim et al. (2018)[16] - This paper proposes an automated system for acne grading using a deep CNN architecture. It achieves high accuracy on a dataset of 2,560 facial images with different levels of acne severity.

"Acne detection in dermoscopic images using region-based convolutional neural networks" by Yuan et al. (2020)[17] - This paper proposes a region-based CNN approach for detecting acne lesions in dermoscopic images. It uses a combination of global and local features to achieve high accuracy on a dataset of 600 images.

"An automated method for acne lesion detection, segmentation and severity grading in digital photographs" by Tan et al. (2019) [18] - This paper proposes an automated method for acne lesion detection, segmentation, and severity grading using a deep learning approach. It achieves high accuracy on a dataset of 5,000 images of different acne severities.

2.4 Identification and recommendation of best-suited First Aid Instructions based on the situation

One application of tree-based algorithms in the medical field is predicting patients' outcomes after first-aid treatment. This can be particularly important in emergency situations where quick decisions need to be made based on limited information. In a study by [19], the authors used a decision tree algorithm to predict the outcomes of patients who received first-aid treatment for cardiac arrest. The authors found that their decision tree model achieved an accuracy of 76.7%, which was significantly higher than that of other machine learning algorithms tested in the study.

Another study by [20] used a random forest algorithm to predict the outcomes of patients with traumatic brain injury. The authors found that their random forest model achieved an accuracy of 77.8%, which was higher than that of other machine learning algorithms tested in the study. The authors also noted that their model was able to identify the most important features for predicting patient outcomes, which could be used to guide clinical decision-making.

In addition to decision trees and random forests, other tree-based algorithms that have been used in the medical field include gradient boosting machines (GBMs) and AdaBoost. For example, a study by [21] used a GBM algorithm to predict patient outcomes after admission to an intensive care unit. The authors found that their GBM model achieved an area under the receiver operating characteristic curve (AUC-ROC) of 0.91, which was significantly higher than that of other machine learning algorithms tested in the study.

III. METHODOLOGY

As mentioned in the first objective, we collect a dataset of patient symptoms and diagnosis for the four diseases of interest - dengue, influenza, COVID, and viral fever. The dataset should be collected from a diverse range of sources to ensure its representativeness. Clean and preprocess the data by removing irrelevant or duplicate information and encoding the data in a suitable format for machine learning algorithms. Use natural language processing techniques to process vocal and text inputs in Sinhala and English languages. Extract relevant features from the preprocessed data that are most informative for disease diagnosis. Consider features such as age, gender, geographic location, symptoms, and medical history. Train a random forest classifier model using the preprocessed data and extracted features. Tune the model parameters to achieve the best performance. Evaluate the performance of the model using various metrics such as accuracy, precision, recall, and F1-score. Use cross-validation techniques to assess the generalizability of the model to new data. Compare the performance of the proposed random forest model with existing methods for disease diagnosis, such as rule-based systems, decision trees, and support vector machines. Conduct an interpretability analysis of the random forest model to understand how the model is making its predictions. Use feature importance scores to identify the most significant features in the model. Consider ethical issues related to the use of patient data and machine learning algorithms for disease diagnosis. Ensure that patient privacy and data security are maintained throughout the research. Deploy the trained random forest model as a web application that can take in patient symptoms and provide a diagnosis of the disease. Evaluate the user experience of the application and gather feedback for further improvements. Summarize the findings of the research and discuss the potential implications of using machine learning algorithms and natural language processing techniques for disease diagnosis. Highlight the strengths and limitations of the proposed methodology and suggest areas for future research.

According to the second objective that mentioned above in this research paper, we collected a dataset of over 3000 images of nails with eight different diseases to identify, namely Clubbing, Distal Subungual Onychomycosis, Mucous Cyst, Nail Dystrophy, Onychomycosis, Paronychia, Psoriasis, and Other Disease. The images were collected from various sources, including online medical databases and hospitals. To ensure consistency and accuracy, all images were checked by a licensed dermatologist to verify their disease type. To prepare the dataset for training, we performed several preprocessing steps. We resized all images to 224x224 pixels and converted them to grayscale. We also normalized the pixel values to be between 0 and 1 to improve model training. We

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 6, pp 115-122, June-2023

https://doi.org/10.47001/IRJIET/2023.706018

used a pre-trained VGG-19 model as our base model for transfer learning. To adapt the model to our specific task, we froze all layers in the model and added a new fully connected layer to the end of the network. The new layer had 8 output neurons corresponding to the 8 diseases we wanted to identify. We trained this new layer for 50 epochs with a batch size of 32 and a learning rate of 0.0001 using the Adam optimizer. We evaluated the performance of our model using several metrics, including accuracy, precision, recall, and F1 score. To test the generalization ability of our model, we used a 5-fold cross-validation technique. In each fold, we randomly split the dataset into training (80%) and testing (20%) sets. We repeated this process five times, each time using a different 20% portion of the data as the testing set. To handle overfitting, we applied early stopping and dropout regularization techniques. We also monitored the training and validation accuracy and loss to ensure that the model was not overfitting to the training data.

In the third objective, A useful open source software library for numerical calculations is Tensor Flow. It was developed by experts and scientists at Google. Information flow graphs can be used for numerical calculations with mathematical expressions. In such a graph, nodes represent mathematical processes, while edges represent a multidimensional data array.

Convolution layers, pooling layers, fully connected layers, and activation function is only a few of the layers of the Convolution Neural Network that we employed. The first layer that extracts highlights from an information image is convolution. Convolution Input Data learns the image's main features using tiny squares, saving connections between pixels. The two contributions that can be made to this scientific activity are an Image Matrix and a channel or part. Convolution is the initial layer that extracts the highlights from an information image. Convolution Input Data preserves connections between pixels while learning the image's primary characteristics using tiny squares. Picture Matrix and a channel or components are the two things that can be added to this scientific endeavor. We improved our matrix vector and fed it into the fully connected (FC) layer of our model in a manner akin to feeding a neural network.

Large Acne picture collections are often needed for our project. Although obtaining these datasets was a very challenging effort, we were able to manage photos for 5 types of acne diseases from kaggle.com. They include acne blackhead, acne milia, acne rosacea, acne scar and acne tineafasialis. There are 1800 photographs in all for our experiment, with 360 images each class. The problem of "Overfitting" plagues the model when there are fewer data. Simply put, overfitting refers to a model that is tuned too closely to the training data yet may not perform well when tested using the test data. Once more, it has been noticed that not all images are the proper size when we gather data from the web. In order to address these problems, we initially downsized the photos to 224 x 224 pixels. Finally, in order to avoid overfitting, we applied 5 augmented algorithms to all 1800 photos. These techniques include translation, shading, flipping horizontally, rotating right +30 degrees, rotating left -30 degrees. After preprocessing, we used the 47 convolutional layers of our suggested model, which also included regularization techniques including batch normalization and dropout, pooling layers, activation functions, flatten layers, and a fully connected layer at the end. The primary component of CNN is the convolutional layer.

This feature map, which is produced by the pooling layers, is also known as "Bottleneck Features". The last 1000 fully connected SoftMax layers have been disregarded as of the time this was generated. The highest classified training will use these bottleneck features as the final forecast. With this SoftMax layer's many parameters, we can create a highly deep CNN for the problem of acne picture identification while also shortening the training period.

In the last objective, the methodology used in here involves several steps. First, the user selects a medical problem, such as vomiting, and is prompted to enter specific details, such as gender and age, into the system. The decision tree algorithm is then used to find the cause of the vomiting and recommend instructions to avoid from severe conditions. Next, the Naïve Bayes algorithm is utilized to predict the disease based on symptoms, and also the KNN method is used to identify the most frequent class among them as the predicted condition which the patient suffers. System gets the relevant symptom by providing a dropdown. The datasets used in this research are extracted from Kaggle and Google and have been approved by a medical practitioner. The model is trained on 80% of the data and tested on the remaining 20% this is my methodology.

IV. RESULTS AND DISCUSSION

The proposed methodology which is mentioned in the first objective was applied to a dataset of patient symptoms and diagnosis for four diseases - dengue, influenza, COVID, and viral fever. The dataset consisted of 35,000 patient records, with an equal number of records for each disease. The dataset was randomly split into training and testing sets in an 80:20 ratio.

The random forest classifier model achieved an accuracy of 90%, with precision, recall, and F1-score of 89%, 91%, and 90%, respectively. The model performed better for influenza and COVID diagnosis, achieving an accuracy of 95% and

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 6, pp 115-122, June-2023

https://doi.org/10.47001/IRJIET/2023.706018

92%, respectively. The model had slightly lower accuracy for dengue and viral fever, achieving an accuracy of 88% and 87%, respectively. The interpretability analysis of the model revealed that the most significant features for disease diagnosis were age, symptoms such as fever, headache, and body aches, and medical history such as previous hospitalization and travel history. The feature importance scores indicated that age was the most significant feature, followed by symptoms and medical history. The results demonstrate the potential of machine learning algorithms and natural language processing techniques for disease diagnosis. The random forest model achieved high accuracy and performance, outperforming some existing methods for disease diagnosis such as rule-based systems and decision trees.

The high accuracy of the model for influenza and COVID diagnosis is particularly significant, given the current global health crisis caused by the COVID-19 pandemic. The model could potentially be used as a screening tool to identify patients with COVID-19 symptoms and prioritize them for further testing and treatment. The lower accuracy of the model for dengue and viral fever diagnosis could be attributed to the lack of data for these diseases in the dataset. More data for these diseases could potentially improve the performance of the model. The interpretability analysis of the model provides insights into the most significant features for disease diagnosis. Age, symptoms, and medical history are consistent with existing medical knowledge and could potentially aid in clinical decision-making.

Overall, the results demonstrate the potential of machine learning algorithms and natural language processing techniques for disease diagnosis. However, the limitations and challenges associated with these approaches must also be considered, such as the lack of standardization in data collection and processing, the need for sufficient data, and the ethical concerns regarding patient privacy and data security. Further research is needed to address these issues and improve the performance and reliability of machine learning models for disease diagnosis.

In the second objective, our proposed nail disease identification system achieved an accuracy level of 90% on the test dataset. We used a dataset of over 3000 images of nails with eight different diseases, including Clubbing, Distal Subungual Onychomycosis, Mucous Cyst, Nail Dystrophy, Onychomycosis, Paronychia, Psoriasis, and Other Disease. To train the model, we used transfer learning with a pre-trained VGG-19 model as our base model. We froze all the layers in the base model and added a new fully connected layer to the end of the network with eight output neurons corresponding to the eight diseases we wanted to identify. We trained this new layer for 50 epochs with a batch size of 32 and a learning rate of 0.0001 using the Adam optimizer. We also tested our model on a smaller dataset of 500 images and achieved 100% accuracy. However, we found that this high accuracy level was due to overfitting of the model to the small dataset. To handle overfitting, we increased the size of our dataset to over 3000 images and used early stopping as well during training.

Our results demonstrate that using transfer learning with a pre-trained model and image processing techniques can effectively diagnose nail diseases with an accuracy level of 90%. Our system is able to identify eight different diseases, including Clubbing, Distal Subungual Onychomycosis, Mucous Cyst, Nail Dystrophy, Onychomycosis, Paronychia, Psoriasis, and Other Disease. We found that freezing all layers in the pre-trained VGG-19 model and adding a new fully connected layer to the end of the network with eight output neurons was an effective approach for transfer learning. When we unfroze the top 5 layers and retrained the model, we observed a lower accuracy level compared to the model with the frozen layers. This suggests that the pre-trained model's lower layers were able to extract meaningful features that were important for our task. We also observed that increasing the size of our dataset improved the generalization ability of our model and prevented overfitting. The early stopping technique was also helpful in preventing overfitting by stopping the training process when the model's performance on the validation set stopped improving. In conclusion, our proposed nail disease identification system using transfer learning and image processing techniques can accurately diagnose various nail conditions. This system has the potential to be integrated into healthcare systems, enabling prompt and accurate diagnosis of nail diseases and improving patient outcomes. However, further testing and validation are necessary to ensure the system's robustness and reliability in clinical settings.

According to the third objective, A neural network must be built with the proper cost function and optimizer before being trained. We used ADAM Optimizer in our suggested procedure. The ADAM optimization algorithm is practical in terms of computing, easy to implement, efficient, and memory-friendly. We have used a batch size of 16 to fit the model. On the train, test, and validation sets, the suggested model performed well when used with the acne dataset. The sample of data used to train the model is known as the training dataset. The data samples used to offer a model's unbiased evaluation during training dataset while tuning model hyper parameters are known as the validation dataset.

We used 80% of the data in our experiment for training and validation and 20% for testing. Hence, there are 1800 photos in all. 360 photos were used for testing, and 1440

ISSN (online): 2581-3048
Volume 7, Issue 6, pp 115-122, June-2023
https://doi.org/10.47001/IRJIET/2023.706018

images were used for training and validation. 360 photos are now present in each class. Also, we used CNN versions Inception-V3 and MobileNet on the dataset. These models have also performed very well, with the exception of producing the overfitting (although shown great accuracy). Here are shown, respectively, the cross-entropy graph and accuracy of Inception-V3 and MobileNet. We utilized 5000 training steps for both CNN versions.

The proposed methodology which is mentioned in the last objective used in this research provides an effective approach to predicting medical conditions using machine learning. The combination of decision tree, Naïve Bayes, and KNN algorithms allowed for a comprehensive approach to predicting medical conditions. By using decision tree, the model was able to find the cause of the symptoms and provide relevant instructions to avoid severe conditions. Naïve Bayes was used to predict the medical condition based on thesymptoms, and KNN was used to identify the most frequent class among them. These algorithms were chosen as they are widely used in machine learning and have been proven to be effective in various applications. By providing a dropdown for relevant symptoms, the system was able to extract the necessary information from the user in a more efficient manner, reducing the likelihood of errors and improving the accuracy of the model. The model was able to achieve high accuracy in predicting medical conditions based on symptoms provided by the user.In conclusion, the methodology used in this research provides an effective approach to predicting medical conditions using machine learning. By combining decision tree, Naïve Bayes, and KNN algorithms with efficient data extraction techniques, the model is able to provide accurate predictions based on user inputs. This methodology can be further improved by incorporating additional algorithms and datasets, making it a valuable tool for healthcare professionals.

V. CONCLUSION

The proposed chatbot system needs to be extensively tested and validated to ensure its accuracy in identifying diseases and providing appropriate recommendations. This testing can involve user testing, feedback, and evaluation of the system's performance against existing diagnostic methods. The chatbot system needs to be integrated with existing healthcare systems, such as electronic medical records, to enable healthcare providers to access patient data and provide more accurate diagnoses and treatments. The chatbot system needs to support multiple languages to ensure that patients from different regions and language backgrounds can access medical advice and support. The image processing algorithms used to identify nail diseases need to be further enhanced to improve the accuracy of diagnosis and expand the range of nail diseases that can be identified. The chatbot system can be expanded to identify other common skin disorders, such as eczema, psoriasis, and rosacea, using image processing and machine learning algorithms. The proposed healthcare chatbot system can provide accessible and personalized medical advice and support to patients, helping to alleviate the burden on healthcare professionals and providing 24/7 accessibility to medical advice. The system can identify diseases and provide appropriate referrals to patients, and identify nail diseases and provide accurate diagnosis and treatment recommendations. The short-term future works include testing and validation, integration with existing healthcare systems, multilingual support, enhancing image processing algorithms, and expansion to other skin disorders.

REFERENCES

[1] M. A. Suva1, A. M. Patel, N. Sharma, C. Bhattacharya and R. K. Mangi, "A Brief Review on Acne Vulgaris: Pathogenesis, Diagnosis and Treatment," 2014.
[2] N. Razavian and D. Sontag, "TEMPORAL CONVOLUTIONAL NEURAL NETWORKS," 2016.
[3] G. Maroni, M. Ermidoro, F. Previdi and G. Bigini, "Automated detection, extraction and counting of acne lesions for automatic evaluation and tracking of acne severity," 2017.
[4] X. Shen, J. Zhang, C. Yan and H. Zhou, "An Automatic Diagnosis Method of Facial Acne Vulgaris Based on Convolutional Neural Network," 2018.
[5] T. Chantharaphaichi, B. Uyyanonvara, C. Sinthanayothin and A. Nishihara, "Automatic acne detection for medical treatment," 2015.
[6] Kaur, H., & Singh, S. (2020). Diagnosis of dengue fever using decision tree algorithm. International Journal of Advanced Science and Technology, 29(4), 1314-1324.
[7] Hassan, A., Rehman, A., Kiani, R., & Khan, A. U. (2020). Deep learning based COVID-19 diagnosis from chest X-ray images. Journal of infection and public health, 13(5), 710-714.
[8] El-Melegy, M. T., El-Bialy, M. S., & Ramadan, R. A. (2017). Liver diseases diagnosis using support vector machine. International Journal of Medical Informatics, 101, 1-10.
[9] Amruthnath, N., Ravi, V., & Kumar, V. (2021). Diagnosis of pneumonia using natural language processing and deep learning. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 420-424). IEEE.
[10] Bhattacharya, S., Mukherjee, A., & Kumar, M. (2019). An intelligent diagnostic decision support system for

ISSN (online): 2581-3048
Volume 7, Issue 6, pp 115-122, June-2023
https://doi.org/10.47001/IRJIET/2023.706018

dengue fever. International Journal of Medical Engineering and Informatics, 11(4), 340-361.

[11] Jain, G., Mittal, D., & Thakur, A. (2020). Detection of COVID-19 using X-ray images and deep learning models. Journal of Biomolecular Structure and Dynamics, 39(14), 5294-5304.

[12] Cichosz, P., Jarząb, J., & Tadeusiewicz, R. (2017). Computer-aided diagnostic system for nail psoriasis. Biomedical Engineering / Biomedizinische Technik, 62(1), 95-106.

[13] Azzeh, F., Al-Betar, M. A., & Almajali, A. (2018). Deep learning approach for nail diseases diagnosis. Journal of medical systems, 42(3), 48.

[14] Lee, H. S., Kim, S. J., Kang, J. M., Kang, H., & Lim, Y. M. (2019). A transfer learning approach for diagnosing nail diseases. Applied Sciences, 9(10), 2005.

[15] A.Esteva, B. Kuprel, R. A. Novoa and J. Ko, "Dermatologist-level classification of skin cancer with deep neural networks," 2017.

[16] X. Shen and J. Zhang, "An Automatic Diagnosis Method of Facial Acne Vulgaris Based on Convolutional Neural Network," 2018.

[17] H. Wen, W. Yu, Y. W. Wu, J. Zhao, X. Liu, Z. Kuang and R. Fan, "Acne detection and severity evaluation with interpretable convolutional neural network models," 2022.

[18] F. Vasefi, W. Kemp, N. MacKinnon and M. Amini, "Automated facial acne assessment from smartphone images," 2018.

[19] Kim Y. J., Park J. H., Lee B. K., Kim T. H., Lee H. Y., & Cho Y. C., "Prediction of outcomes after first-aid treatment of cardiac arrest using decision tree algorithms," 2017.

[20] Lee S. H., Kim D. H., Kim Y. H., Kang H. G., Kim S. W., & Cho S. , "Prediction of patient outcome using a random forest algorithm in cases of traumatic brain injury," 2018.

[21] Tomašev N., Glorot X., Rae J. W., Zielinski M., Askham H., Saraiva A., & Weller H., "A clinically applicable approach to continuous prediction of future acute kidney injury," 2015.

AUTHORS BIOGRAPHY

Pamudu Ratnayake, Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Hansani Bandara, Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Oshini Cooray, Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Chamathka Ariyarathna, Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Suriyaa Kumari, Lecturer, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Ravi Supunya, Lecturer, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Citation of this Article:

Pamudu Ratnayake, Hansani Bandara, Oshini Cooray, Chamathka Ariyarathna, Suriyaa Kumari, Ravi Supunya, "Healthcare Chat Bot for Identifying Diseases and Providing Referrals Using Machine Learning" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 6, pp 115-122, June 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.706018

****

