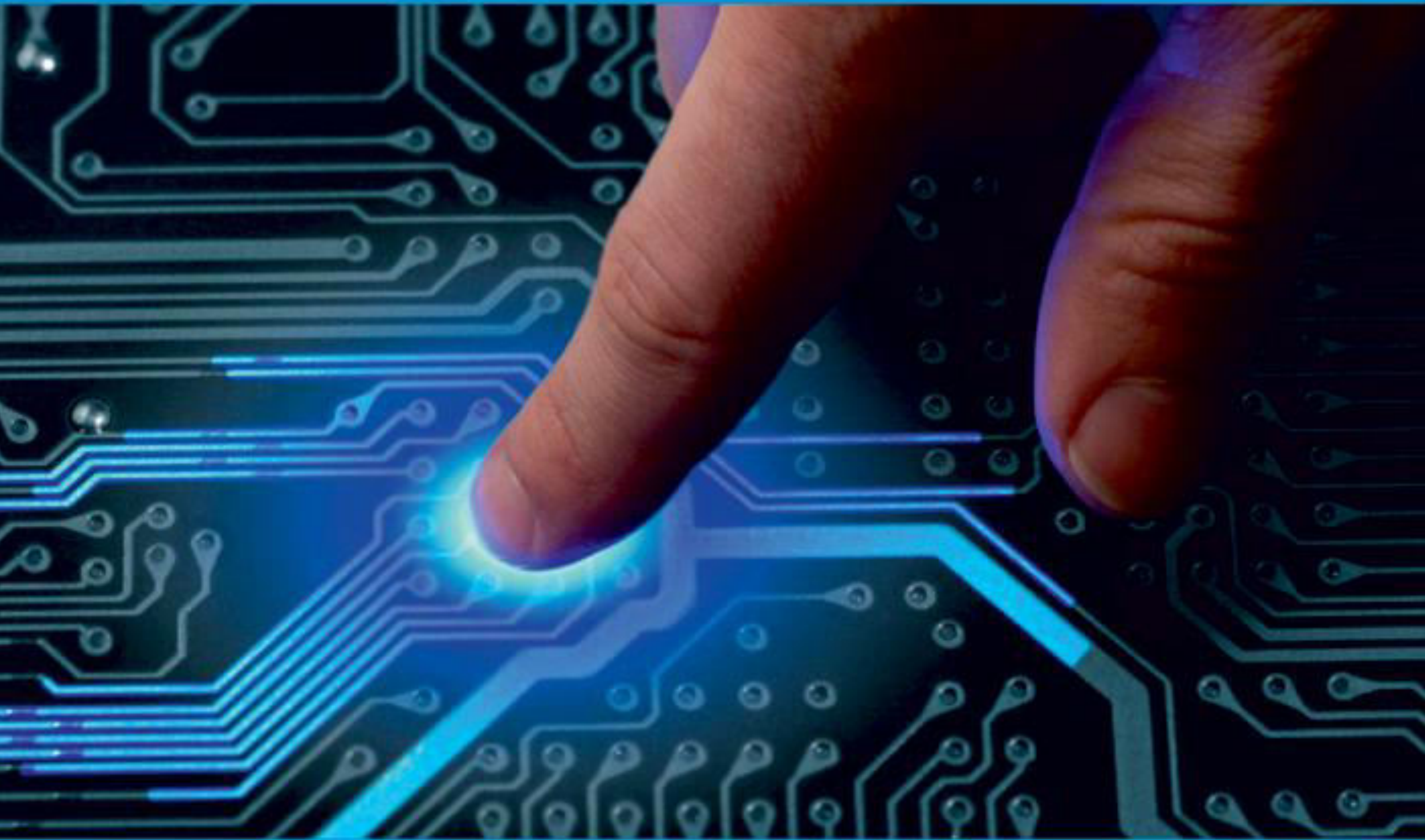
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 9, September 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

9940 572 462 6381 907 438 ijircce@gmail.com www.ijircce.com

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 9, September 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1009014 |

# Healthcare Predication Using ML

**V.C.Bagal , R.S.Shetty**

Assistant Professor, K.K.Wagh Institute of Engineering Education and Research, Hirabai Haridas Vidyanagri, Amruthdham , Panchavati, Nashik- Maharashtra, India

Student, K.K.Wagh Institute of Engineering Education and Research, Hirabai Haridas Vidyanagri, Amruthdham ,Panchavati, Nashik- Maharashtra, India

**ABSTRACT**: Analyzing health records to avert future complications and provide the right treatment plays an important role in medical diagnostics. This paper introduces a method to use real life ICD-coded Electronic Medical Records (EMR), clinical data collected between 2018-2019 in Africa, to create a prediction model. The prediction model uses recent advances in machine learning, specifically Convolutional Neural Networks (CNN), to provide an alternative and a powerful way for disease risk prediction. An algorithm is first designed to process the real-life EMR data into multi-step time-series forecasting data. A CNN model is then developed, to predict the individual's future disease class (ICD-10 CM chapters) risk based on their demographic and medical history. The experimental results show that the model predicts future disease risk, for an individual, with an accuracy of 80.73%. We conclude that such a model can prove immensely useful in cost savings for individuals and hospitals by performing preemptive corrective action, provide additional guidelines for hospitals' capacity planning, when the model is applied for demography, and, most importantly, peace of mind on personal health.

KEYWORDS**:** Electronic Medical Records, Machine Learning, Convolutional Neural Network, Disease Prediction, Data Processing.

## I. INTRODUCTION

Healthcare prediction using patient treatment history and health data by applying data mining and machine learning techniques is ongoing struggle for the past decades. Many works have been applied data mining techniques to pathological data or medical profiles for prediction of specific diseases. These approaches tried to predict the reoccurrence of disease. Also, some approaches try to do prediction on control and progression of disease. The recent success of deep learning in disparate areas of machine learning has driven a shift towards machine learning models that can learn rich, hierarchical representations of raw data with little preprocessing and produce more accurate results. With the development of big data technology, more attention has been paid to disease prediction from the perspective of big data analysis; various researches have been conducted by selecting the characteristics automatically from a large number of data to improve the accuracy of risk classification rather than the previously selected characteristics. Also, the cost incurred in storing data in centralized servers is more and many times users have to pay for the entire plan which they have selected even if they have used only a fraction of storage portion thus it does not provide flexibility to the user to payonly for what they are using. Another issue is the scalability of the system, it is difficultto scale a centralized storage system to meet the increasing demand. The main focus is on to use machine learning in healthcare to supplement patient care for better results. Machine learning has made easier to identify different diseases and diagnosis correctly. Predictive analysis with the help of efficient multiple machine learning algorithms helps to predict the disease more correctly and help treat patients. The healthcare industry produces large amounts of healthcare data daily that can be used to extract information for predicting disease that can happen to a patient in future while using the treatment history and health data. This hidden information in the healthcare data will be later used for affective decision making for patient's health. Also, these areas need improvement by using the informative data in healthcare.

## II. LITERATURE SURVEY

D. A. Davis, N. V. Chawla, N. Blumm, et al., has designed system that can assist a medical practitioner in decision making. If a sampling of future diagnoses can be provided to a practitioner, appropriate medical tests can be ordered sooner and lifestyle adjustments can be adopted by the patient proactively. This will not only result in improving the quality of life for the patient, but also in reducing the health care costs. To that end, we proposed CARE, a collaborative recommendation engine for prospective and proactive healthcare. CARE relied solely on the ICD disease

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 9, September 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1009014 |

codes, which are a standard across insurance and medicare databases. This exploitation of ICD codes by CARE allows for a seamless integration with a variety of electronic healthcare systems that use or will embrace the standard of ICD. Also, as the medical community moves toward comprehensive electronic records, CARE becomes increasingly relevant.[1]

Amisha, P. Malik, M. Pathania, et al., proposes that AI promises to change the practice of medicine in hitherto unknown ways, but many of its practical applications are still in their infancy and need to be explored and developed better. Medical professionals also need to understand and acclimatize themselves with these advances for better healthcare delivery to the masses.[2]

M. M. Rahman, B. C. Desai, and P. Bhattacharyapresented a content-based image retrieval framework for diverse collections of medical images of different modalities, anatomical regions, acquisition views, and biological systems. For the image representation, the probabilistic output from multi-class support vector machines (SVMs) with low-level features as inputs are represented as a vector of confidence or membership scores of pre-defined image categories. The outputs are combined for feature-level fusion and retrieval based on the combination rules that are derived by following Bayes' theorem. We also propose an adaptive similarity fusion approach based on a linear combination of individual feature level similarities. The feature weights are calculated by considering both the precision and the rank order information of top retrieved relevant images as predicted by SVMs. The weights are dynamically updated by the system for each individual search to produce effective results. The experiments and analysis of the results are based on a diverse medical image collection of 11,000 images of 116 categories. The performances of the classification and retrieval algorithms are evaluated both in terms of error rate and precision–recall. Our results demonstrate the effectiveness of the proposed framework as compared to the commonly used approaches based on low-level feature descriptors.[3]

## III. DATASET


Fig.a: Covid Dataset (Covid)

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 9, September 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1009014 |


Fig. A1 Normal

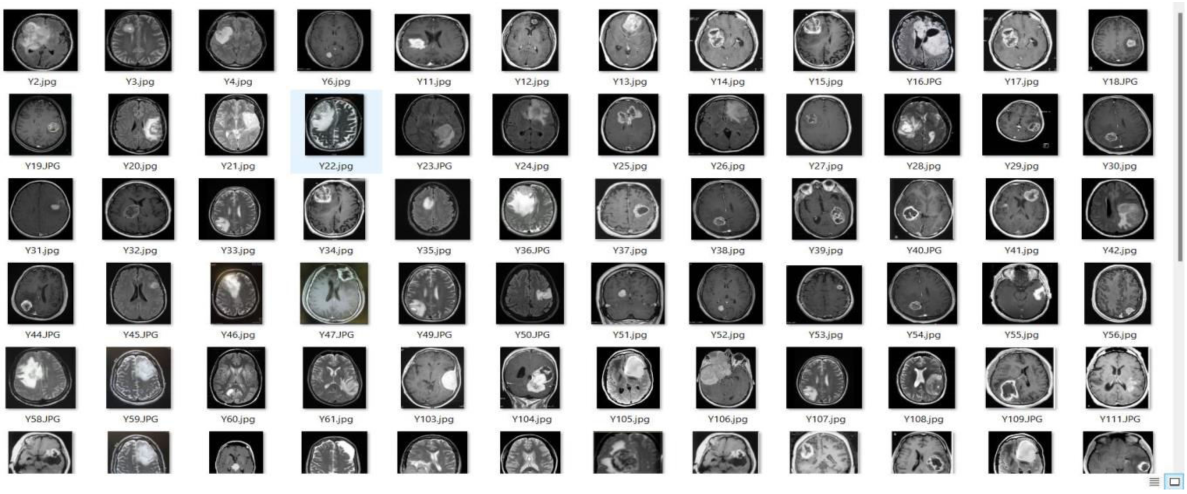
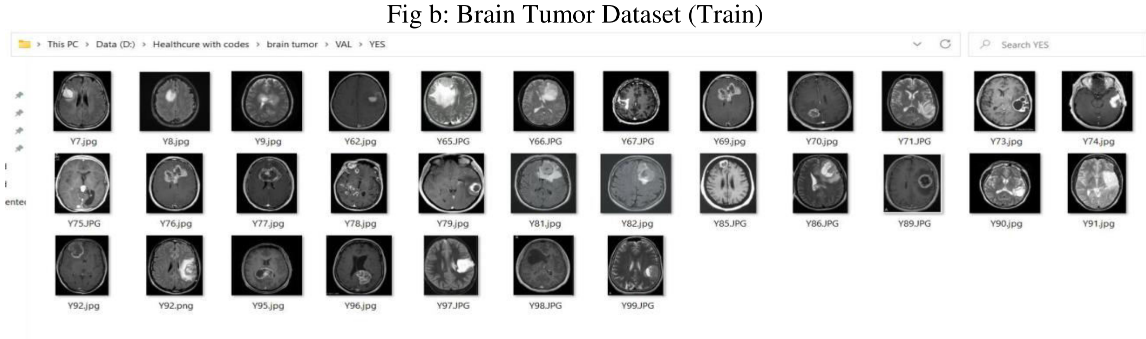



Fig b: Brain Tumor Dataset (Train)

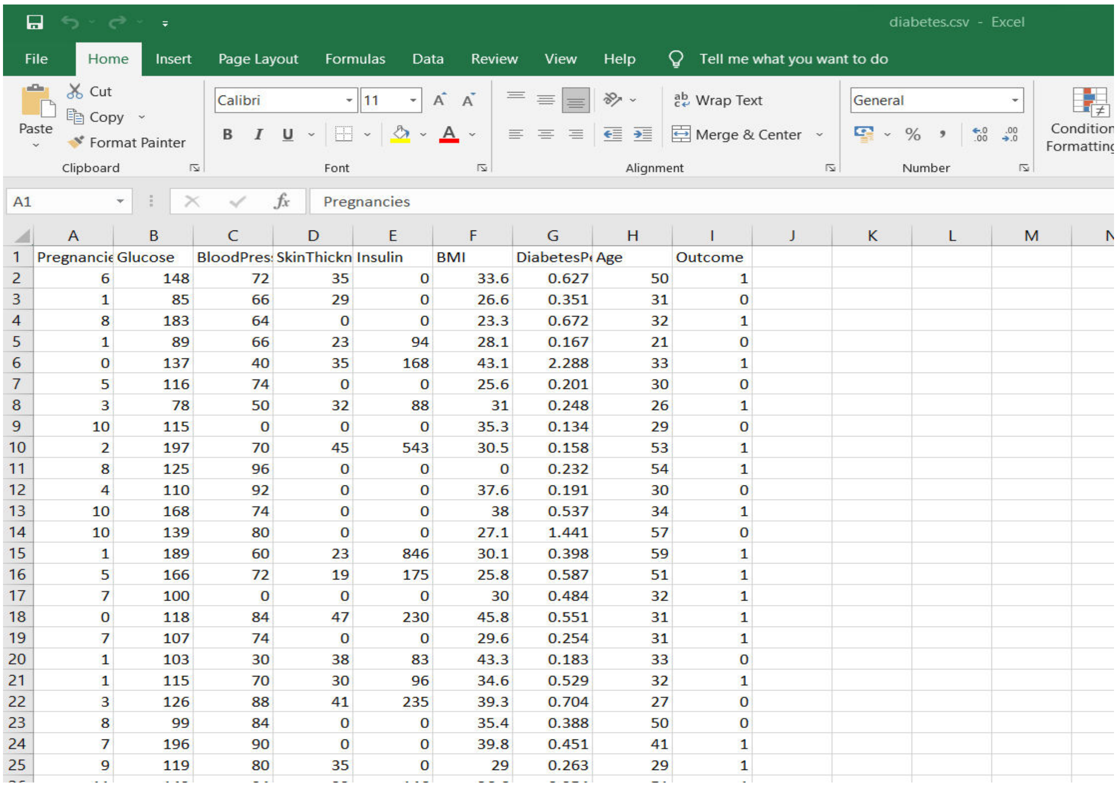



Validation Dataset (Test)

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 9, September 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1009014 |



| Pregnancie | Glucose | BloodPres | SkinThickn | Insulin | BMI | DiabetesP | Age | Outcome |
|---|---|---|---|---|---|---|---|---|
| 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 | 1 |
| 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |
| 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | 0 |
| 3 | 78 | 50 | 32 | 88 | 31 | 0.248 | 26 | 1 |
| 10 | 115 | 0 | 0 | 0 | 35.3 | 0.134 | 29 | 0 |
| 2 | 197 | 70 | 45 | 543 | 30.5 | 0.158 | 53 | 1 |
| 8 | 125 | 96 | 0 | 0 | 0 | 0.232 | 54 | 1 |
| 4 | 110 | 92 | 0 | 0 | 37.6 | 0.191 | 30 | 0 |
| 10 | 168 | 74 | 0 | 0 | 38 | 0.537 | 34 | 1 |
| 10 | 139 | 80 | 0 | 0 | 27.1 | 1.441 | 57 | 0 |
| 1 | 189 | 60 | 23 | 846 | 30.1 | 0.398 | 59 | 1 |
| 5 | 166 | 72 | 19 | 175 | 25.8 | 0.587 | 51 | 1 |
| 7 | 100 | 0 | 0 | 0 | 30 | 0.484 | 32 | 1 |
| 0 | 118 | 84 | 47 | 230 | 45.8 | 0.551 | 31 | 1 |
| 7 | 107 | 74 | 0 | 0 | 29.6 | 0.254 | 31 | 1 |
| 1 | 103 | 30 | 38 | 83 | 43.3 | 0.183 | 33 | 0 |
| 1 | 115 | 70 | 30 | 96 | 34.6 | 0.529 | 32 | 1 |
| 3 | 126 | 88 | 41 | 235 | 39.3 | 0.704 | 27 | 0 |
| 8 | 99 | 84 | 0 | 0 | 35.4 | 0.388 | 50 | 0 |
| 7 | 196 | 90 | 0 | 0 | 39.8 | 0.451 | 41 | 1 |
| 9 | 119 | 80 | 35 | 0 | 29 | 0.263 | 29 | 1 |

Fig C: Diabetes dataset


Fig C: Pneumonia dataset (Normal)

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 9, September 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1009014 |

Fig.C1 :Pneumonia dataset

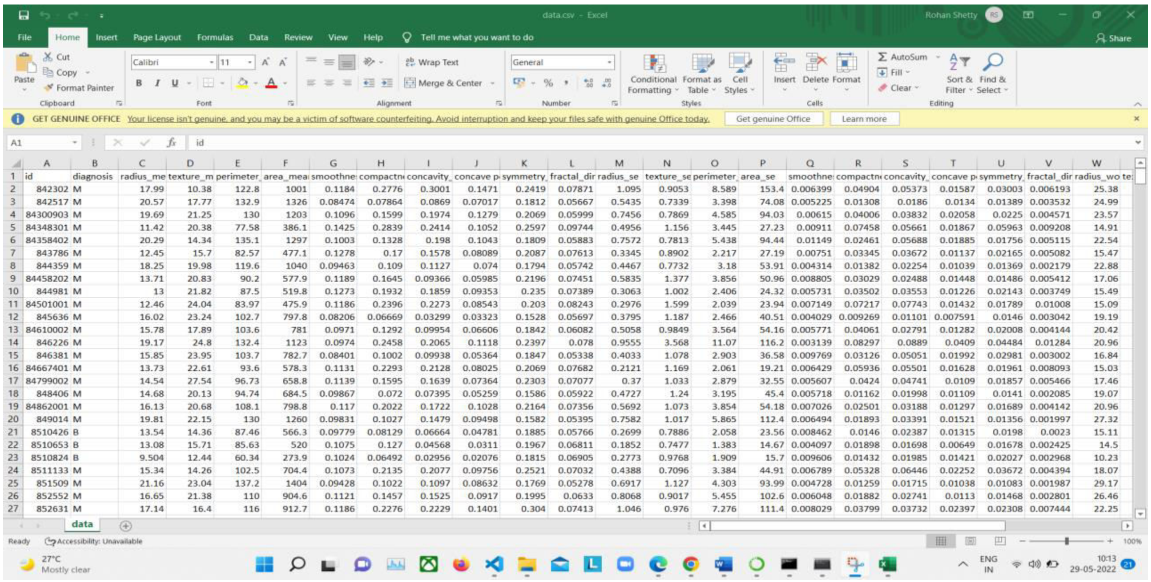


Fig D: Breast Cancer Dataset

## IV. METHODOLOGY

Detection of disease at prodigal stage is very important as it can prevent serious damage to the patient's brain. The detection of the disease is done the medical professionals throughcertain steps such as CT scan. Thedetection of the disease is time consuming using CT scan. Hence, we are applying the machine learning technology to predict the disease. The proposed approach detects the various stages of Disease such as moderate-demented and non-dementedusing CNN algorithm. It reduces the time required to predict the output and can be used for real time predictions. We describethe datasets used in this study and how the data was pre-processed before the machinelearning task. Feature extraction using principal component analysis and feature selection techniques were also employed. After the data preprocessing is done, the efficient machine learning algorithm that is CNNis applied to predict the disease and categorized it into moderate demented and non demented.

1] Data preprocessing Datapreprocessing refers to all transformations on the raw data before it is fed to the machine learning or deep learning algorithm. For instance, training a convolutional neural network on raw images will probably lead to bad classification performances. The preprocessing is also important to speed uptraining such as clustering and scaling technique. Real- world data is often incomplete, inconsistent, and lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is aproven method of resolving such issues.

2] Feature selection and extraction Feature selection is an effective strategy to optimize the predictive performance of machine learning algorithms. In this study, we have tested a number of feature selection techniques including

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 9, September 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1009014 |

SelectKBest. Sequential Forward Selection Sequential Backward Selection, and Recursive FeatureElimination implemented in the sklearn package. The models with the best performance resulted from the SelectKbest, which is a univariate correlation feature selection method. We explored values for different number of features (k) with the highest correlation to the dependent variable (AD/control).

3] Convolution Neural Network Artificial Neural Networks are used in various classification task like image, audio, words. Different types of Neural Networks are used for different purposes, for example for predicting the sequence of words we use Recurrent Neural Networks more precisely an LSTM, similarly for image classification we use Convolution Neural Network. In the next section, we are going to build basic building block for CNN. In a regular Neural Network, there are three types layers. Input Layer: It's the layer in which we give input to our model. The number of neurons in this layer is equal to total number of features in our data (number of pixels in case of an image). Hidden Layer: The input from Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data. Each hidden layers can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by addition of learnable biases followed by activation function which makes the network nonlinear. Output Layer: The output from the 14 hidden layer then fed into a logistic function like sigmoid or softmax which converts the output of each class into probability score of each class. The data is then fed into the model and output from each layer is obtained this step is called feed forward, we then calculate the error using an error function, some common error functions are cross entropy, square loss error etc. After that, we back propagate into the model by calculating the derivatives. This step is called back propagation which basically is used to minimize the loss. The implementation was carried out using Python programming language with Scikit- learn, Pandas, NumPy, and Matplotlib libraries using CNN algorithm.

## V. CONCLUSION

Medical records provide an important historical perspective into a individual's health. Capturing that knowledge to find general patterns is incredibly useful to sustain a healthy population, since it provides possibilities for preventive action. Advancements in machine learning offer an interesting alternative to traditional prediction models developed in the medical field. In this work, an algorithm is developed to convert a real-life EMR ICD-10 CM coded data into a multi-step time-series forecasting data for each individual. This chronological information is then used to train a CNN model to predict the future disease risk class. The model achieved an accuracy of 80.73% in correctly predicting the disease risk class, while this can be improved further by incorporating more information about the individual, like the history of hereditary disease in the family. The results achieved show that machine learning models can provide a strong alternative to traditional methods. It would also be interesting to compare, the model's disease prediction with 93 a medical practitioner's perspective. For all the positives, it should also be noted that such a model can be misused/abused to take advantage of the vulnerable people.

## REFERENCES

[1] D. A. Davis, N. V. Chawla, N. Blumm, et al., "Predicting individual disease risk based on medical history," in Proceeding of the 17th ACM conference on Information and knowledge mining - CIKM '08, New York, New York, USA: ACM Press, 2008, p. 769, ISBN: 9781595939913. DOI: 10.1145/1458082. 1458185.

[2] Amisha, P. Malik, M. Pathania, et al., "Overview of artificial intelligence in medicine," Journal of family medicine and primary care, vol. 8, no. 7, pp. 2328–2331, 2019, ISSN: 2249-4863. DOI: 10 . 4103 / jfmpc .jfmpc{\textunderscore } 440{\textunderscore}19.

[3] M. M. Rahman, B. C. Desai, and P. Bhattacharya, "Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," Computerized Medical Imaging and Graphics, vol. 32, no. 2, pp. 95-108, 2008.

[4] S. Hussain, S. M. Anwar, and M. Majid, "Segmentation of Glioma Tumors in Brain Using Deep Convolutional Neural Network," arXiv preprint arXiv:1708.00377, 2017. [5] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1207-1216, 2016.

[6] P. Moeskops et al., "Automatic segmentation of MR brain images with a convolutional neural network," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1252-1261, 2016

[7] M. Havaei et al., "Brain tumor segmentation with deep neural networks," Medical image analysis, vol. 35, pp. 18-31, 2017

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 9, September 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1009014 |

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details