International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 6, Nov-Dec-2024, ISSN (Online): 2395-566X

Diabetes Prediction Using Neural Network

Anand Singh, Vedant Urkudkar, Ruchi vairagade, Ketaki Punjabi
Department of Computer Science Engineering,
MIT Art, Design and Technology University, Pune, India

Abstract- Diabetes is one of the most frequent diseases worldwide where yet no remedy is discovered for it. Every year a great deal of money has to be spent for caring for patients with diabetes. Therefore, it is crucial that prediction should be very accurate and a very dependable method must be adopted for doing so. One of these methods is the use of artificial intelligence systems, and in particular, the use of Artificial Neural Networks, or ANN. So, in this paper, we used artificial neural networks in order to predict whether or not a person has diabetes. The criterion was to minimize the error function in neural network training with the help of a neural network model. After training the ANN model, the average error function of the neural network was equal to 0.01 and the accuracy of the prediction of whether a person is diabetics or not was 70%

Index Terms- diabetes, neural network, ANN, prediction.

I. INTRODUCTION

Diabetes is a chronic condition that occurs when the pancreas cannot produce enough insulin or when the body cannot use its own insulin. Insulin is a hormone produced by the pancreas that regulates the level of sugar in the blood. Hyperglycemia or hyperglycemia is one of the most common effects of uncontrolled diabetes and, after some time, causes serious damage to many organs, notably nerves and blood vessels.

In 2015, 8.5% of adults aged 17 years or older had diabetes. In 2013, diabetes was the cause of 1.5 million deaths, and high blood glucose caused 2.3 million deaths. Diabetes patients have doubled in the last ten years all over the world. More than 200 million people are infected and about seven percent increase in the annual predominance of diabetes in the world. People for a long period suffered from various illnesses and in some cases have been able to identify their diseases and provide them with the remedy in order to improve it, but sadly, at times due to the lack of symptom diagnosis in patients for a long period may even endanger the life of the patient. Therefore, many studies have been done in the field of predicting for several diseases to the extent that today's human take advantage of decision supports models and smart method to predict. One of the decision support models application is in the medical field and diagnosis of illnesses such as diabetes [1, 2]. Deferment in the diagnosis and prediction of diabetes due to insufficient control of blood glucose increases macro vascular and Capillaries difficulties risk, ocular diseases and kidney failure [1, 2]. So we propose an ANN model that may predict diabetes useful and helpful for doctors and practitioners.

Some of the attributes used in this research are: Number of pregnancies, PG Concentration (Plasma glucose at 2 hours in an oral glucose tolerance test), Diastolic BP (Diastolic Blood Pressure (mm Hg))), Tri Fold Thick (Triceps Skin Fold Thickness (mm)), Serum Ins(2-Hour Serum Insulin (mu U/ml)), BMI (Body Mass Index: (weight in kg/ (height in m)^2)), DP Function(Diabetes Pedigree Function), Age (years), Diabetes (Whether or not the person has diabetes)[15]. According to reports of the Diabetes Research Center, incidence of diabetes folded in the last ten years worldwide and more than 200 million people are infected and about seven percent increase in the annual prevalence of diabetes worldwide.

Since diabetes is a long-term disease that import permanent damage to the limbs and vital organs in the body, the use of artificial intelligence tools helps enhance the detection methods and disease control which will be a great help to the physicians. According to the Diabetes Research Center, it has been illustrated that early diagnosis of patients at risk can prevent up to 80 percent of lasting complications of type II diabetes or deferred them [5].

There are two types of diabetes: type I and type II diabetes. Type I diabetes also termed insulin dependent and type II diabetes termed relative insulin deficiency [6]. The protracted complications of diabetes are mainly classified under two categories: vascular and nonvascular complications of diabetes.

Vascular complications include micro vascular (eye disease, neuropathy, nephropathy) and macro vascular complications (coronary artery disease, peripheral vascular disease, cerebrovascular disease). Non-vascular complications include gastro paresis, sexual dysfunction, and skin changes [7].

The Objectives of the Study

- To predict and categorize the state of health.
- To identify some suitable factors that impact health conditions,
- To develop an artificial neural network that can be used to predict health performance based on defined information about a specific health condition

II. LITERATURE REVIEW

- To forecast and classify the health status.
- To determine some relevant factors that influence health status,
- To develop an artificial neural network that may be applied to predict health performance given some specific predefined data for a specific health status

Diabetes or diabetes mellitus is a form of metabolic disorder (metabolic) of the body. This disease destroys the ability to produce insulin in the patient's body or the body develops resistance to insulin the and consequently the produced insulin cannot achieve its normal job. The main role of the produced insulin is to decrees blood sugar by different instruments. There are two key types of diabetes. In type I diabetes, the destruction of pancreatic cells ruins insulin synthesis and, in type II, there's a gradual insulin resistance within the body and eventually can lead to the destruction of pancreatic beta cells and alterations in insulin secretion. In type II diabetes, one knows that genetic defects, obesity and sedentary lifestyle play an important role in a human being [1]. Even though the exact cause of type I diabetes is unknown, factors that might suggest a higher predisposition include the following:

- Family history. The risk increases for a person if his parent or sibling has history of having type I diabetes.
- Environmental factors. Conditions such as exposure to a viral infection most likely play some role in type I diabetes.

The presence of destructive immune system cells. Sometimes relatives of a patient with type I diabetes are tested for the presence of diabetes autoantibodies. If a person has these autoantibodies, he /she has the opportunity of higher risk for developing type I diabetes. However not every individual who has those autoantibodies develops diabetes.

- Geography. Some nations, such as Sweden, have higher incidences of type I diabetes.

Researchers don't fully understand why certain people develop pre-diabetes and type II diabetes and others don't. It's sure that some factors upsurge the risk like [2]:

Weight. The more fatty tissue you have, the more resilient a person cells to insulin.

Inactivity. The less energetic a person is, the more a person has risk. Physical activity assists a person control of his/her weight, consumes glucose as energy and makes a person cells more sensitive to insulin.

Family history. A person risk upsurges if his parent or sibling has history of type II diabetes.

Race. Even though it's uncertain why, people of specific races are at higher risk.

Age. The risk of a person upsurges as he/she grows older. This may be caused by a fact that a person has a habit to exercise less, lose muscle mass and add weight as he/she grows older. Nevertheless type II diabetes is also rising among children, youths and adults.

Gestational diabetes. If a person developed gestational diabetes when she was pregnant, her risk of emerging pre-diabetes and type II diabetes far ahead upsurges. If she gives birth to a baby weighing more than 4 kilograms, she is also at risk of type II diabetes.

Polycystic ovary syndrome. For females, having polycystic
High blood pressure. Having blood pressure more than 140/90 millimeters of mercury (mm Hg) is connected to an augmented risk of type II diabetes.

The imbalances of cholesterol and triglycerides. One who's at low level of good cholesterol or high-density lipoprotein is likely to have a risk of type II diabetes. In addition to this, the third kind of fat in the body is passed through the blood. More are the levels of triglycerides the more the risk of having type II diabetes.

A practical approach to this type of problem is application of regression analysis where past data is better combined into some functions. The result is an equation in which both xj inputs are multiplied by wj; the sum of all these products is constant and then output $y = \Sigma$ wj xj +, where j = 0.n.

The problem is that choosing an appropriate function is difficult so that all the collected data is to be adjusted within it such that the output can be adjusted automatically when more information is attained, since the candidate's performance is organized by a number of arguments, and this control will not have any clear regression model.

This type of problems can be addressed through a more common approach-the artificial neural network, which emulates human thinking while trying to solve the problem. Thus, the attempt to develop an adaptive system such as artificial neural network to predict the situation and classification based on the results of these arguments.

1. Artificial Neural Network

Adaptive Artificial Neural Network is a non-parametric technique to categorize that in the medical field based on input variables to categorize subjects into healthy or unhealthy. Classification and prediction of the patient's condition based on risk factors are an application of artificial neural networks. Furthermore, ANN is an application of Artificial Intelligence.

In artificial neural networks is inspired by the diverse structure of the human brain. Billions of nerve cells (neurons) through the communication that with each other (synapses) creates a biological neural network in the human brain that is devoted to human activities like speaking, reading, comprehension, breathing, face detection, movement, voice recognition, also resolve issues and data storage. Artificial neural networks, in fact, mimic a part of brain jobs.

2. Artificial Neural Network Structure

- Neural networks are nonlinear modeling of intelligent computational methods that, recently, is considered to be one of the advance tools in computing and information processing, and during acquiring a high position in the science field, the consequences were quite promising. Feedforward neural networks are a precious type of artificial neural networks, as feedforward neural network with a hidden layer and the appropriate activation function in the hidden layer with sufficient neurons in the hidden layer can approximate any function with any accuracy. For this purpose, in the next section we introduce a structure for the feedforward neural network modelling to the prediction diabetes problem.
- Generally, the artificial neural network has three types of layers.
- Input layer: Obtains the raw data that has been fed to the network.
- tHidden layers: their functions are determined by the inputs, weight, relationship between them, and the hidden layers. It is based on the weights between input and hidden units whether a hidden unit needs to be activated or not.
- Output layer: output unit function based on the activity and weight of the hidden unit and the connectivity between the hidden units and the output.

3. The Back-propagation Training Algorithm

- Set each wi to some small random value
- Until the stopping condition is satisfied, Do
- For each training example <(x1,…xn),t> Do
- Apply the instance (x1,…,xn) to the network and calculate the network output values ok
- For each output unit k:

□k=ok(1-ok)(tk-ok)

- For each hidden unit h:

□h=oh(1-oh) □k wh,k □k

•For each network weight wj Do

•wi,j=wi,j+□wi,j,where □wi,j= □ □j xi,j and □ is the learning rate.

4. Previous Studies

The author in [16] made use of Data Mining to come up with a model that classifies diabetic patient control level based on historical medical records. The death caused by diabetes in the world necessitated avoiding complications brought about by the disease. Based on this, he came up with a new predictive model that was developed using data mining techniques, which could classify diabetic patient control level based on historical medical records. The work was undertaken with the help of three data mining methods which are Naïve Bayes, Logistic and J48. The work was set out using the WEKA application. The result presented that Logistic data mining algorithm produced a precision average of 0.73, recall of 0.744, Fmeasure of 0.653 and accuracy of 74.4%. Naïve Bayes produced a precision average of 0.717, recall of 0.742, F-measure of 0.653, and accuracy of 74.2%. J48 produced a precision average of 0.54, recall of 0.735, F-measure of 0.623 and accuracy of 73.5%. This signifies that the logistic algorithm was more accurate than the other two. The research, however, was limited as it only considered diabetes type 2. They also failed to look into the discovery of appropriate features with minimal effort and validation on discovered features. •\tThe author in [17] used data mining to develop a prediction model for diabetes Type II treatment plans.

He was encouraged by the highly dangerous complication of chronic disease as well as the complication which required amputation of one of the parties. He developed a new model for classifying diabetes type 2 treatment plans that could help the control of blood glucose level of diabetic patient. He used the J48 algorithm in carrying out the experiment on 318 medical records collected from JABER ABN ABU ALIZ clinic centre for diabetes in Sudan. The general control data indicated that 59.1% of the record was taken into consideration as Oral Hypoglycemic, 35.5% for Insulin and 5.3% for Diet. He carried out the evaluation using the WEKA application. The research work did not encompass diabetes type 1 patients which could have been included with additional attributes. Moreover, the nutrition system and exercise could have been added to enhance the precision of the system. • The authors in [18] used prediction of diabetes mellitus based on boosting ensemble modeling. They were motivated by the focus of aiding diabetes patients fit themselves into their normal activities of life by early predicting their state and tacking it.

They intended to predict the types of diabetes patients would have based on physical and clinical information by using the boosting ensemble technique. They used boosting ensemble

technique which internally utilizes a random committee classifier. The architecture utilized was supported by integrating data management, learning, and prediction components together. The result of the evaluation of the technique with accuracy gave a weighted average TP rate of 0.81, FP rate of 0.198, Precision of 0.81, Recall of 0.81, F-measure of 0.82 and ROC area of 0.82 for diabetes type 1 and 2. The research work is intended to be extended in future the integration into a cloud based clinical decision support system for chronic diseases and the inclusion of a feedback mechanism to increase the level of satisfaction of users. •\tSernyak used logistic regression analysis to calculate odds ratio neuroleptic unusual version and a diagnosis of diabetes in each of the age groups, control the effects of population, and diagnosis [9]. Thirugnanam has improved diabetes prediction using fuzzy neural networks [10]. Hamid and others have offered hybrid intelligent systems for the detection of micro albuminuria in patients with type 2 diabetes without measuring the urinary albumin [11].

- Javad and others proposed the method base on automatic learning on type II diabetes to regulate blood sugar.

III. METHODOLOGY

By contemplating seriously through literature and seeking experience from human experts on pathological conditions, various factors have been seen to impact the determination of patients' cases in the subsequent period. These factors were carefully studied and coordinated appropriately with a number suitable for coding the computer inside the modeling environment ANN. These factors were classified as input variables and output variables that indicate some feasible levels of disease status based on the assessment system. The data were entered to the JNN tool environment, calculated the value of each of the variables based on JNN(the strongest influential factor on diabetes), then the data were trained, validated, and tested.

1. Input Variables

The specified input variables are those that can be obtained simply from the file system and the registry of diseases. Input variables are:

Table 1: attributes in the Data set

No.	Attribute name
1	Pregnancies: Number of pregnancies
2	PG Concentration: Plasma glucose at 2 hours in an oral glucose tolerance test
3	Diastolic BP: Diastolic Blood Pressure (mm Hg)
4	Tri Fold Thick: Triceps Skin Fold Thickness (mm)
5	Serum Ins: 2-Hour Serum Insulin (mu U/ml)
6	BMI: Body Mass Index: (weight in kg/ (height in m)^2)
7	DP Function: Diabetes Pedigree Function
8	Age: Age (years)
9	Diabetes: Whether or not the person diabetes

These factors were converted into a format suitable for neural network analysis as shown in Table2 "data set up to 1004", Input characteristics 8 and one output (0 diabetic, 1 healthy)

2. The Output Variable

The output variable represents whether a person has diabetes or not (Sick, Healthy).

Table 3: Output Data Transformation

S/N	Output Variable	Diabetes
1	Healthy "1 "	The person does not have diabetes
2	Sick "0"	The person has diabetes

Table 3 shows the classification of the selected output variable, which is consistent with the classification system, in the identification of disease cases.

3. Neural Network Evaluation

As stated earlier, the goal of this experiment is to determine whether the individual has diabetes or not. We used the Backpropagation algorithm, which gives the capability of learning and testing a neural network. Our neural network is the feed front network with 1 input layer of 8 inputs, 3 hidden layers, and 1 output layer of 1 output as it appears in Figure 2. The proposed model is implemented in Just Neural Network (JNN) environment. The dataset for the diagnoses of diabetes were gathered from the documentation of the Association of diabetic's city of Urmia which contains 1004 samples with 9 attributes (as seen in Fig 1). This model was used to deduce the value of every one of the variables using JNN which they are the most influential factor on diabetes prediction as shown in figure 3. Upon training and validating, the network, it was tested by use of test data and the following results were obtained. The accuracy of the diabetes predication was (87.3%). The average error was 0.010. The training cycles (number of epochs) were 158,000. The training examples were 767. The number of validating examples was 237 as seen in figure 4.

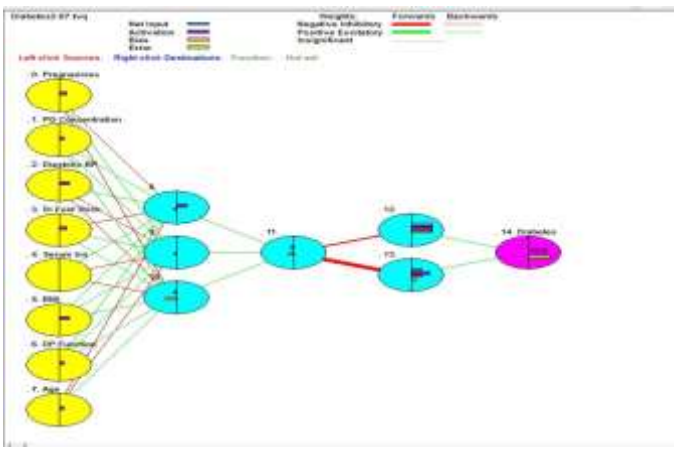

Figure 1: Artificial Neural Network Structure

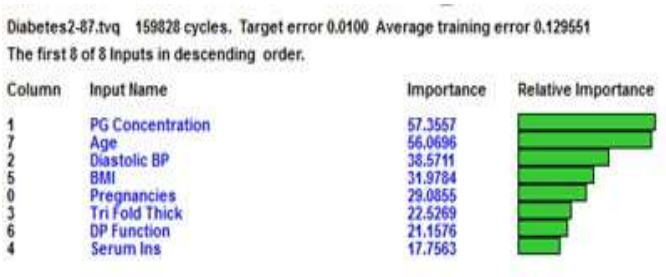

Figure 2: Attributes importance (the most influential factor on diabetes)

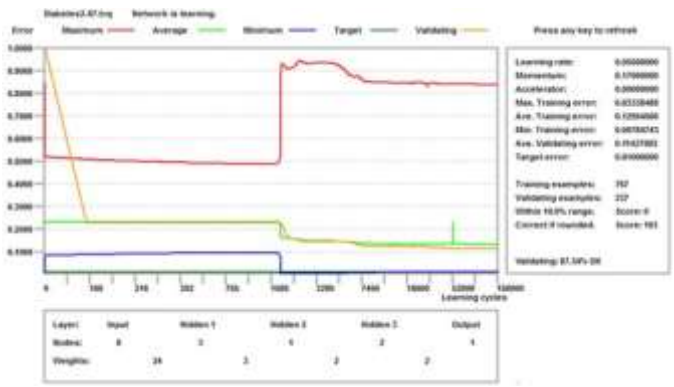

Figure 3: Learning progress

IV. CONCLUSION

In this paper, artificial neural network was used to predict diabetes. By using an artificial neural networks model we can design and implement complex medical processes using software. The more effective and efficient the system is in various medical fields ranging from predicting to their problems, diagnosing them and treating and helping the surgeons, physicians and the general population. These systems can be implemented in a parallel way and are distributed in various measures. Artificial neural network is a parallel processing system in general, which is used to detect complex patterns within the data. In this study, the aim was to find the effective variables and determine how they would affect diabetes. A proposed model was implemented in JNN environment. This model had 1004 samples with 9 attributes, used to find out the value of each one of the variables using JNN (the most influencing factor on diabetes) upon training, validating, and testing the given dataset we are able to get an accuracy of (87.3%), average error was (0.010), number of epochs was (158,000), number of training examples was (767), and number of validating examples was (237).

REFERENCES

1. Priyanka Paygude, Milind Gayakwad, Dhanashri Wategaonkar, Rajendra Pawar, Ramchandra Pujeri, Rahul Joshi, Dried fish dataset for Indian seafood: A machine learning application, Data in Brief, Volume 55,2024, 110563, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2024.110563.
2. Kuldeep Vayadande, Amol A. Bhosle, Rajendra G. Pawar, Deepali J. Joshi, Preeti A. Bailke, Om Lohade, Innovative approaches for skin disease identification in machine learning: A comprehensive study, Oral Oncology Reports, Volume 10, 2024, 100365, ISSN 2772-9060, https://doi.org/10.1016/j.oor.2024.100365.
3. P. M. Shelke, R. G. Pawar, P. Pawar, R. Rathod and T. Kawoor, "Synergizing Investment Portfolios: A Technologically Advanced Approach with Integrated Chatbot Capabilities," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-6, doi: 10.1109/MITADTSoCiCon60330.2024.10575311.
4. D. Shah, A. K. Singh, A. Prasad, S. Koppikar, B. Dixit and R. G. Pawar, "MIT-WPU Insight Sync: A Data Empowerment Platform," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-5, doi: 10.1109/MITADTSoCiCon60330.2024.10575731.
5. R. Pawar, S. Ghumbre and R. Deshmukh, "Visual similarity using convolution neural network over textual similarity in content- based recommender system", International Journal of Advanced Science and Technology, vol. 27, pp. 137-147, Sep. 2019.
6. Pawar, Rajendra, Shashikant Ghumbre, and Ratnadeep Deshmukh. "Developing an Improvised E-Menu Recommendation System for Customer." Recent Findings in Intelligent Computing Techniques: Proceedings of the 5th ICACNI 2017, Volume 2. Springer Singapore, 2018.
7. Dixit, B. ., G. Pawar, R. ., Gayakwad, M. ., Joshi, R. ., Mahajan, A. ., & Chinchmalatpure, 6Suyash V. . (2023). Challenges and a Novel Approach for Image Captioning Using Neural Network and Searching Techniques. International Journal of Intelligent Systems and Applications in Engineering, 11(3), 712–720.
8. Garg, A. et al. (2023). Slider-Crank Four-Bar Mechanism-Based Ornithopter: Design and Simulation. In: Tuba, M., Akashe, S., Joshi, A. (eds) ICT Systems and Sustainability. Lecture Notes in Networks and Systems, vol 516. Springer, Singapore. https://doi.org/10.1007/978-981-19-5221-0_26
9. Dhanashree Wategaonkar, Rajendra Pawar, Prathamesh Jadhav, Tanaya Patole, Rohit R. Jadhav and Saisrijan Gupta, "SIGN GESTURE INTERPRETER FOR BETTER COMMUNICATION BETWEEN A NORMAL AND DEAF PERSON", Journal of Pharmaceutical Negative Results, pp. 5990-6000, 2022.
10. R. Pawar, S. Ghumbre and R. Deshmukh, "A Hybrid Approach towards Improving Performance of

Recommender System Using Matrix Factorization Techniques", International Journal of Future Generation Communication and Networking, vol. 13, no. 4, pp. 467-477, 2020

11. V. Sinha, S. Chinchmalatpure, S. Apte and R. Pawar, "NeuroSight: Brain Tumor Classification using Deep Learning Techniques," 2023 International Conference on Artificial Intelligence Science and Applications in Industry and Society (CAISAIS), Galala, Egypt, 2023, pp. 1-6, doi: 10.1109/CAISAIS59399.2023.10270090.
12. S. Kumar, O. Parhad, D. Pathak, S. Irlewad, P. Parira and R. Pawar, "Forecasting the Rate of Crimes Committed Against Women Using Machine Learning Approach and Data Mining Techniques," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-5, doi: 10.1109/ICCUBEA58933.2023.10392097.
13. A.Patil, R. Khatavkar, S. Kardile, P. Kakade and R. Pawar, "Computer Controlled Letter Writing Robotic Arm Using Speech Recognition," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-7, doi: 10.1109/ICCUBEA58933.2023.10392038.
14. S. Tyagi, K. Panchal, P. Chitta, S. Todi, S. Priya and R. Pawar, "Emotionomics: Pioneering Depression Detection Through Facial Expression Analytics," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-4, doi: 1109/ICCUBEA58933.2023.10392157.
15. T. Bachal, A. Kumar, A. Shetty and R. Pawar, "Analyzing Weather Patterns and Mosquito Data to Predict West Nile Virus Outbreaks in Chicago," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-5, doi: 10.1109/ICCUBEA58933.2023.10392129.
16. R. Pawar et al., "Crop Advancement with Machine Learning," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-6, doi: 10.1109/ICCUBEA58933.2023.10392151.
17. A.Kulkarni, P. Bapat, T. Kulkarni and R. Pawar, "Review of Reinforcement Learning in Chrome Dino Game," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-5, doi: 10.1109/ICCUBEA58933.2023.10392095.
18. A.Kelkar, D. Jondhale, R. Borade, I. Juneja, Y. Haran and R. G. Pawar, "Revolutionizing the Driving Experience : Enhancing Safety and Comfort Using ADAS," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-4, doi: 10.1109/ICCUBEA58933.2023.10392267.
19. A.Markad, S. Phadke, P. Shah and R. Pawar, "Real-Time CCTV Face Recognition Model," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-4, doi: 10.1109/ICCUBEA58933.2023.10392020.
20. Padmanabh Butala, Dr. Rajendra Pawar, Dr. Nagesh Jadhav, Manas Kalangan, Aniket Dhumal, & Sahil Kakad. (2024). Deep Learning Approach towards Emotion Recognition Based on Speech. JOURNAL OF ADVANCED APPLIED SCIENTIFIC RESEARCH, 6(3). Retrieved from http://mail.joaasr.com/index.php/joaasr/article/view/948
21. Priyanka Paygude, Milind Gayakwad, Dhanashri Wategaonkar, Rajendra Pawar, Ramchandra Pujeri, Rahul Joshi, Dried fish dataset for Indian seafood: A machine learning application, Data in Brief, Volume 55,2024, 110563, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2024.110563
22. P. M. Shelke, R. G. Pawar, P. Pawar, R. Rathod and T. Kawoor, "Synergizing Investment Portfolios: A Technologically Advanced Approach with Integrated Chatbot Capabilities," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-6, doi: 10.1109/MITADTSoCiCon60330.2024.10575311.
23. R. Pawar, S. Ghumbre and R. Deshmukh, "A Hybrid Approach towards Improving Performance of Recommender System Using Matrix Factorization Techniques", International Journal of Future Generation Communication and Networking, vol. 13, no. 4, pp. 467-477, 2020.