

Diabetes Diagnostic Method based on Tongue Image Classification Using Machine Learning Algorithms

M. Bharathi¹, Dr.D. Prasad², Dr.T. Venkatakrishnamoorthy³, Dr.M. Dharani⁴

¹Assistant Professor, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupthi, Andhrapradesh.

²Professor, Sasi Institute of Technology & Engineering, Tadepalligudem, Andhrapradesh.

³Associate Professor, Sasi Institute of Technology & Engineering, Tadepalligudem, Andhrapradesh.

⁴Associate Professor, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupthi, Andhrapradesh.

Abstract

Machine learning algorithms are implemented in variety of medicine applications in real time systems. In this sector, image classification plays a main role. Present days many people are suffering from Diabetes Mellitus, which is a disease caused by high blood glucose levels in the blood. Having too much glucose in your blood can leads to diabetes over time. Diabetes Mellitus also takes a toll on the entire human body, which effects on oral health. In real time, there are several algorithms are implemented for classification of various types of diseases for obtain good accuracy levels in classification levels. Artificial Intelligence shows major role in medicine applications in image classification. As a result, many techniques based on Machine Learning and Deep Learning have been developed. In this process, the classification can be performed with diabetes tongue, which preprocess for, extract features by various standard classification algorithms such as K-Nearest Neighbors, Support Vector Machine, Decision Tree, Logistic Regression, Convolution Neural Networks. The proposed algorithm developed with dimensional reduction-based CNN process, which is more accurate compared with other classification algorithms.

Keywords: Diabetes Mellitus, Deep Learning, K-Nearest Neighbors, Convolution Neural Networks, Support Vector Machine.

DOI:10.47750/pnr.2022.13.04.174

INTRODUCTION

In the world, Diabetes is one of the deadliest diseases which affects more and more individuals on a daily basis. The general identifying process is that patients consult diagnostic center or doctor to spend more time to get their reports, it is more time consuming and money to get their diagnosis report. Therefore, how to quickly and accurately diagnose and analyze diabetes is a topic worth studying [1] [2]. The machine learning algorithm help people make a preliminary judgment about diabetes mellitus using tongue images, and it can serve as a reference for doctors. Blood glucose, also referred to as blood sugar, is excessively high in people with diabetes can be monitored regularly [18]. The glucose remains in the bloodstream and does not reach the cells. As a result, your body is unable to transport sugar or glucose from your bloodstream to your cells, resulting in an overabundance in your bloodstream [13]. An excessive amount of glucose in the blood might lead to health concerns over time. Although there is no cure for diabetes, there are things you can do to manage it and stay healthy [3] [4]. In 2020, according to National diabetes statistics report 34.2 million humans in the United States are residing with diabetes. That represents 10.5% of the population. In 1980, 108 million human beings international had diabetes. In 2014 it has risen to 422

million. World Health Organization (WHO) estimated that it 700 million adults have affected worldwide. Type-1 diabetes is an autoimmune attack on the pancreatic cells. It prevents them with insulin and necessitating daily insulin injections. Type-2 diabetes patients can manufacture insulin, but their bodies are resistant to it. By eating a nutritious diet, increasing physical exercise, and controlling stress, people with prediabetes can avoid developing type 2 diabetes [16]. Alaa Tharwat et.al [2020], has suggested algorithm of black box-based support vector machine algorithm with great parameters in terms of the visualization, mathematical and geometrical interpretations for obtaining better accuracy levels. In this work and calculated the accuracy of diabetic tongue using the less mathematical calculations, hence reduces the computational time [5]. Marzia Hoque Tania et.al [2019] implemented machine learning algorithm on automatic Tongue diagnosis (ATD) instead of huge analysis comparing with other standard methods. This paper talks about the advantages, applications, and associated research gaps in standard works on ATD systems [6]. Bob Zhang et.al[2014] perform colour correction by advanced algorithm with multiple geometry features based on areas, distance, measurements and their ratios are retrieved. The decision tree is implemented to classify the objects, which is based on standard Chinese medicine five tongue forms

likes acute and triangles, rectangle, square and circle [7]. Ratchadaporn Kanawong et.al[2012] explore supervised machine learning techniques to investigate new modalities for clinical characterization to construct an automatic classification system for unique color-space- based feature set that can be retrieved from tongue photos of clinical patients[8]. Artificial intelligence plays a significant part in the classification of images used in medicine. Numerous methods based on machine learning and deep learning have been created as a result. [20].

DIMENSIONAL REDUCTION

There are many standard techniques are applying in image processing [21], [22] to improve the quality of the images such as DCT, DTF and Wavelets [9],[10]. The wavelets technique gives effective results compare with other standard methods. In this process, the low and high frequency details are replaced in restoration system. For this type of large prefer the machine learning algorithms [11] can be used in many signal processing applications [19]. Linear discriminate analysis (LDA) is similar to analysis of variance (ANOVA) and regression analysis express one dependent variable as a linear combination of other traits. In discriminant analysis employs continuous independent variables as well as categorical dependent variable. Discriminant correspondence analysis is the analogous technique when dealing with categorical independent variables. ANOVA work with categorical independent variables as well as continuous dependent variable [15]. The principal component analysis (PCA) is also similar as LDA and factor analysis look for linear combinations of variables that best explain the data [14]. LDA approaches based on probability density functions

$p(x|y=0)$ and $p(x|y=1)$ both normal distribution with mean and covariance properties (μ_0, Σ_0) and (μ_1, Σ_1) respectively.

From Bayes solution

$$\frac{1}{\sum_0} \ln \left| \frac{p(x|\mu_0)}{p(x|\mu_1)} \right| > T \tag{1}$$

Where T: Threshold

The resulting classifier is referred to as quadratic discrimination analysis (QDA). The related linear combinations are:

$$w \cdot x > c$$

$$x \sum_0 x = x \sum_1 x \tag{2}$$

$$x \sum_0 \mu_i = x \sum_1 x_i \tag{3}$$

$$w = \Sigma^{-1}(\mu_1 - \mu_0)$$

$$c = \frac{1}{2}((\mu_1 + \mu_0) \cdot w)$$

METHODOLOGY

Principal Component Analysis is an unsupervised learning approach used in machine learning to reduce dimensionality. It is based on a lower statistical process (second order) that uses orthogonal transformation to create a collection of linearly uncorrelated characteristics from a set of observations with similar features, the PC components called transformed features used predictive modeling and data analysis [12] [13]. As illustrated in Figure 1, the methodology is divided into six phases: input image [23] [24], data pre-processing, feature extraction, training features and labels, classification, and analysis of results. The project's flow is depicted in the block diagram below. Here total 500 tongue images are collected and among which 140 are diabetic tongue images and remaining non-diabetic tongue images. The images were collected from various internet sources and some are taken from mobile cameras.

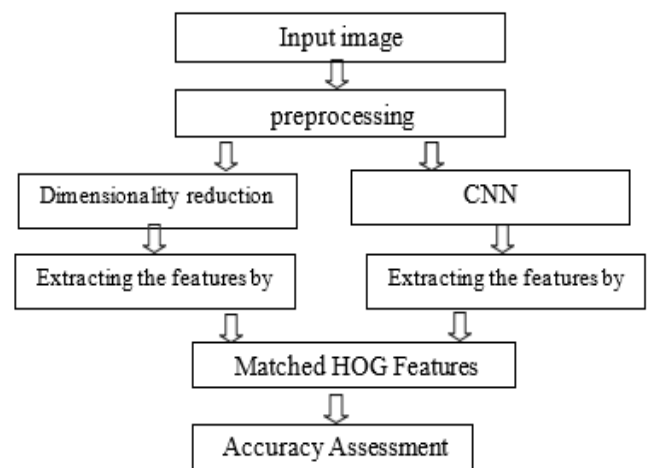


Fig. 3.1: Flow chart for feature extraction

The photos are then categorized as diabetic or non-diabetic, gathered into a folder, and turned into a dataset. The photos are then cropped to the region of interest, which is the upper section of the tongue, and

shrunk to the appropriate size, which is 256x256 pixels. Convolution neural networks have demonstrated exceptional performance in a variety of image classification tasks, including well-known issues. Each convolution layer is made up of a bank of filters (weights) that are combined with the output of the layer below it or, if it's the initial layer, the input image, to form a response [17]. By choosing the highest value in a MxM window, the max-pooling layer subsamples the outputs produced by the previous Convolution layer. The feature extraction procedure is then completed, with hog features used for SVM and KNN and corners used for Decision Tree and Logistic Regression. Then we trained on those traits and assigned labels to them.

RESULT AND DISCUSSION

All of the photos are saved in a file with the labels Diabetic and Non-Diabetic tongue. The images below are examples of tongue imagery.



Fig. 6.1: Non-Diabetic; Fig. 6.2: Non-Diabetic; Fig. 6.3: Diabetic

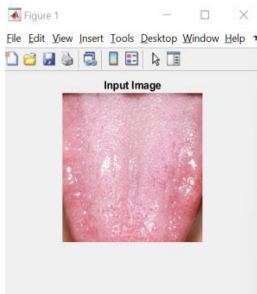


Fig. 6.4. Input Tongue image;

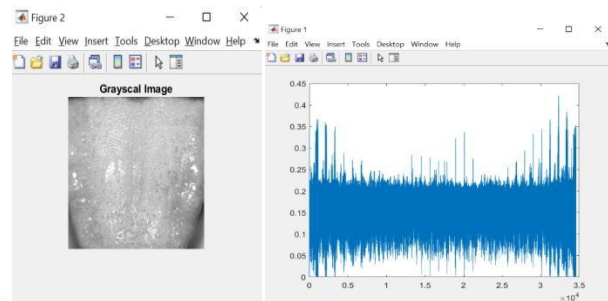


Fig. 6.5. Grayscale image; Fig. 6.6. HOG Features

Then the images are pre-processed as follows first the collected images are cropped to the required region of interest i.e., the upper part of the tongue then the image is resized to (256x256) pixels. The following images show the

image before pre-processing and the image after pre-processing. For SVM, Logistic regression and Decision tree, have changed the images from RGB scale to Grayscale. Then the feature Extraction step is done here have used HOG features and Harris corner features for the different classifiers.

Comparison Table

S.No.	Classifiers	TP	TN	FP	FN	Accuracy
1	KNN (K-Nearest Neighbours)	11	24	01	14	70.00
2	SVM (Support Vector Machine)	18	20	07	05	76.00
3	Decision Tree	19	18	06	07	80.00
4	Logistic Regression	21	19	04	06	74.00
5	CNN (Convolution Neural Network)	23	18	03	05	83.67

Accuracy Formula: $ACC = (TP + TN)/(TP + FP + TN + FN)$ Where TP= True Positive, FP=False Positive, FN=False Negative, TN=True Negative

Here it was analyzed and compared the accuracies of different classifiers decision tree, SVM, KNN Logistic regression and CNN classifier. The accuracy of KNN classifier is recorded as 70%, SVM classifier as 76%, decision tree classifier as 80%, logistic regression classifier as 74% and CNN classifier as 83%. Among these, the proposed algorithm showed better accuracy when compared to remaining classifiers.

CONCLUSIONS & FUTURE SCOPE

A total of 50 tongue photographs were acquired for this experiment, with 25 diabetic images and 25 non-diabetic images. The patient's diagnosis can be made noninvasively based on the features of the tongue image. The accuracy of decision tree, SVM, KNN Logistic regression, and CNN classifiers were studied and compared in this study. KNN classifier has a 70 percent accuracy, SVM classifier has a 76 percent accuracy, decision tree classifier has an 80 percent accuracy, logistic regression classifier has a 74 percent accuracy, and CNN classifier has an 83 percent accuracy. When compared to the remaining classifiers, the CNN classifier performed better. The algorithm can extend the concept that to determine not only positive/negative of tongue but also stages in diabetes. Machine learning algorithm can also be used for other diagnostic purposes like skin cancer detection by suitable machine learning algorithms. The deep learning models, can achieve more and better accuracy when compared to existing works.

REFERENCES

- Pang, B, Zhang, D, LiN, Wang, K, "Computerized tongue diagnosis based on Bayesian networks", *IEEE Trans Biomed Engineering*, 2004, Vol. 51, PP. 803–10.
- Kim, K.H., Do, J.H., Ryu, H., & Kim, J.Y. (2008). Tongue diagnosis method for extraction of effective region and classification of tongue coating, 2008 First Workshops on Image Processing Theory, Tools and Applications (pp. 1-7). *IEEE*.
- N.M. Li, *The contemporary Investigations of Computerized Tongue Diagnosis, The Handbook of Chinese Tongue Diagnosis*, Shed-Yuan Publishing, Beijing, China, 1994.
- Y.G. Wang, J. Yang, Y. Zhou, and Y.Z. Wang, "Region partition and feature matching based color recognition of tongue image," *Pattern Recognition Letters*, 2007. Vol. 28, no. 1, pp. 11–19.
- Alaa Tharwat, Wolfram, Schenck, "Balancing Exploration and Exploitation: A novel active learner for imbalanced data", *Knowledge-Based Systems*, 2020, Volume 210, 106500. [10.1016/j.knsys.2020.106500](https://doi.org/10.1016/j.knsys.2020.106500).
- Marzia Hoque Tania, Khin Lwin, Mohammed Alamgir Hossain, "Advances in automated tongue diagnosis techniques", *Integrative Medicine Research*, 2019. Volume 8, Issue 1, Pages 42-56.
- Bob Zhang, B.V.K. Vijaya Kumar, David Zhang. Detecting diabetes mellitus and Non proliferative diabetic retinopathy using tongue color, Texture, and Geometry features. *IEEE Transactions on Biomedical Engineering*, 61(2), 491-501, 2014.
- Ratchadaporn Kanawong, Tayo Obafemi-Ajayi, Tao Ma, Dong Xu, Shao Li, and Ye Duan, "Automated Tongue Feature Extraction for ZHENG Classification in Traditional Chinese Medicine", *Evidence-Based Complementary and Alternative Medicine*, Hindawi Publishing Corporation, Volume 2012, PP: 1-14.
- D. Leela Rani, M. Bharathi, N. Padmaja, "Performance comparison of FFT,DCT,DWT and DDDWT-OFDM in Rayleigh Channel," *International Conference on Smart Systems and Inventive Technology (ICSSIT 2019)*, Francis Xavier Engineering College, Tirunelveli, India, ISBN Number: 978-1-7281-2118-5, 27th to 29th November, 2019.
- P.L. Lahari, M. Bharathi and Y. Jyothi Shirur, "An Efficient Truncated MAC using Approximate Adders for Image and Video Processing Applications," *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184)*, 2020, pp. 1039-1043, doi: [10.1109/ICOEI48184.2020.9142930](https://doi.org/10.1109/ICOEI48184.2020.9142930).
- Gowthami, B. & Maheswari, C. & Neelima, K. (2017). Face recognition based on SLTP method under different emotions. *International Journal of Electrical and Electronic Engineering and Telecommunications*, 6, 59-66.
- M. Dharani, G. Sreenivasulu, "Land use and land cover change detection by using principal component analysis and morphological operations in remote sensing applications", *International Journal of Computers and Applications*, 2021, Vol. 43, PP. 462-471.
- T. Venkatakrishnamoorthy, GU Reddy, Cloud enhancement of NOAA multispectral images by using independent component analysis and principal component analysis for sustainable systems, 2019, Vol. 74, PP. 35-46.
- Zheng Wang, Feiping Nie, "A New Formulation of Linear Discriminant Analysis for Robust Dimensionality Reduction", May 2018, *IEEE Transactions on Knowledge and Data Engineering PP. (99): 1-13*.
- Debaraj Rana, Swarna Prabha Jena, Subrat Kumar Pradhan, "Performance Comparison of PCA and LDA with Linear Regression and Random Forest for IRIS Flower Classification", *PJAEE*, 2020, Vol. 17(9), PP: 2353-5360.
- H.Z. Zhang; K.Q. Wang; D. Zhang; B. Pang; B. Huang, "Computer Aided Tongue Diagnosis System," *Engineering in Medicine and Biology Society*, 2005. *IEEE-EMBS 2005. 27th Annual International Conference of the Digital Object Identifier*, 2005, pp. 6754 – 6757.
- Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, "Convolutional Neural Networks for Medical Image Analysis: Fine Tuning or Full Training", *IEEE Transactions on Medical Imaging*, 2018, PP. 2018.
- Bharathi, M., Rani, D. L., Padmaja, N., & Dharani, M. (2022). A Health Monitoring System Based on IOT for Persons in Quarantine. *Journal of Algebraic Statistics*, 13(3), 387-392.
- M. Dharani, M. Bharathi (2022). A Novel Realization of Multiplier Design for Signal Processing. *Journal of Pharmaceutical Negative Results* Volume 13, Issue 4, Pages-986-989.
- M. Dharani, Mr. Dhruva Sreenivasa Chakravarthi, Mrs. Vineetha Varghese, Dr. Baig Muntajeb Ali, Dr.K. Venkata Nagendra, Influence of Artificial Intelligence Technology on Teaching Slow Learners, *International Journal of Early Childhood Special Education (INT-JECS)* ISSN: 1308-5581 Vol. 14, Issue 04 2022.
- V. Kalpana, V. Vijaya kishore, K. Praveena, "A Common Frame work for the Extraction of ILD Patterns from CT image", *Publication in Springer Nature in its Lecture Notes in Electrical Engineering (LNEE) Series*, volume 569, ISSN 1876-1100 ISSN 1876-1119 (electronic), ISBN 978-981-13-8941-2 ISBN 978-981-13-8942-9 (eBook). <https://doi.org/10.1007/978-981-13-8942-9>, Dec 2018.
- Praveena, K., Venkatesh, U.S., Sahoo, N.K., Ramanan, S.V., Bee, M.K.M., & Darwante, N. K. (2022). Brain tumor detection using ANFIS classifier and segmentation. *International Journal of Health Sciences*, 6(S3), 11817–11828. <https://doi.org/10.53730/ijhs.v6nS3.8937>.
- R. Pitchai, K. Praveena, P. Murugeswari, Ashok Kumar, M.K. Mariam Bee, Nouf M. Alyami, R.S. Sundaram, B. Srinivas, Lavanya Vadda, T. Prince, "Region Convolutional Neural Network for Brain Tumor Segmentation", *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 8335255, 9 pages, 2022. <https://doi.org/10.1155/2022/8335255>.
- Sakthivel M, Sivanantham S, Kamalraj R & Krishnamoorthy V. An Analysis of Machine Learning Depend on Q-MIND for Defencing the Distributed Denial of Service Attack on Software Defined Network. *International Journal of Early Childhood Special Education*. 2022; 14(05): 3769 – 3776.