**Deep learning model for Tongue Cancer Classification**

**Sajad ABDLKADHIM *[C:\Users\Abdullah\AppData\Local\Microsoft\Windows\INetCache\Content.Word\ORCID-iD_icon-16x16.gif](https://orcid.org/0009-0003-5648-4382),* Sait DEMİR *[C:\Users\Abdullah\AppData\Local\Microsoft\Windows\INetCache\Content.Word\ORCID-iD_icon-16x16.gif](https://orcid.org/0000-0001-8891-4082),* Ashwan A. Abdulmunem *[C:\Users\Abdullah\AppData\Local\Microsoft\Windows\INetCache\Content.Word\ORCID-iD_icon-16x16.gif](https://orcid.org/0000-0002-1903-9269)***

*1 Karabük Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü, Karabük, Türkiye.*

*2* *Karabük Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü, Karabük, Türkiye.*

*3* *University of Kerbala, College of Computer science and information technology, Kerbala, Iraq.*

|  |
| --- |
| **Abstract**  The advancement in computer vision and technology motivated its deployment in various life applications, medical uses of computer vision was one of the main focal of the technology. Image processing with artificial intelligence were the main tool used for medical diagnosis. In this paper, tongue images are used to classify the health. Two classifiers are used namely artificial neural network (FFNN) and Convolutional neural network (CNN). Features extraction also performed using two techniques namely wavelet and image coding. The results show that image coding-based features extraction has optimum results with both FFNN and CNN. |
| Keywords: *Tongue images, Wavelet, Coding, CNN.* |

1. **Introduction**

Image processing technologies have been developed for many years and have been become vital to human daily routine. The population expansion in recent years triggered several challenges about living hood regulation [1]. The advancement of image processing and computer vision has motivated the integration of those techniques in medical diagnosis. It is being ages when medical diagnosed like virus detection in blood samples was performed in time consuming techniques that demand booth funds and man power [2]. So, it is becoming vital to search for alternative which reduce the man power and diagnosing cost. That is obvious in using computer vision based diagnosis system [3].

Microbiology is a field of tracking the unusual creatures inside the biological cells such as tracking virus inside a blood [4].

At [1] Traditional Chinese medicine (TCM) is well known for its use of tongue diagnosis, despite the results of this method's differential diagnosis being occasionally ambiguous and its effects being variable. The modernization of tongue diagnosis led to the development of technology that has standardized and established advantages in clinical settings using CNN.

At [2] The metabolic disorder diabetes mellitus is primarily caused by an increase in blood sugar levels (DM). The most significant health issue of the twenty-first century is swiftly emerging as diabetes mellitus (DM) and the consequences it causes, such as diabetic retinopathy (DR) using UNet-Conditional Random Field-Recurrent Neural Network (UNet-CRF-RNN).

At [4] Some tongue photos can be used for teaching and research in PM diagnosis, which is one of the first steps in standardizing PM diagnostic indices. Nonetheless, more research is necessary to improve the precision of diagnostic models.

Problem statement of this study can be listed in the hereinafter points:

First: detection of such creatures using image processing is troublesome task since the color of virus appears identical with background color which complicate the detection task.

Secondly: from the other hand, detection using deep learning technique e.g. transfer learning is demanding high-end computers with big processing power.

Thirdly, the deployment of virus detection system is required for long term to be integrated with microscopes which is small in size and need a small chip computer.so to say, powerful computers cannot be integrated with such machinery and then using the technology of deep learning in such applications is not possible practically in real life circumstances.

Lastly: The cost of deployment the application in powerful computer is high as compared to standard computer deployment.

1. **METHODOLOGY**

One of the efficient methods to diagnosis the human health is intensive image processing technology. Such technology is applied on the so-called electroluminescence images where the damages amount can be estimated. Three techniques are proposed in the literature to obtain that estimation, firstly analytical approach which tracks the cracks texture and secondly intelligent approach that works on deep learning that intake the image and provide the results without extra added features extraction tasks, lastly hybrid model to estimate cracks and damages is by features extraction integrated with deep/machine learning approach. In hybrid model, aims were to increase the accuracy of damages prediction/classification and reducing the computational costs reported in the first technique (analytical approach).

In this work, two types of features extraction approaches are proposed, using feed forward neural network (FFNN) and secondly using convolutional neural network (CNN). In each technology of above, two types of features extraction are used namely wavelet and features coding. Big data of images are used from open access resource to train the classifiers above. K-fold cross validation is used to obtain the best performance interpretation on the classifiers. Furthermore, in order to get accurate perception about every model performance, K-fold cross validation is to be used. With the mentioned number of database samples, K (folds) can be 10. Thus, accuracy, MAE, Precession can be measured for each fold.

Images are realized in different levels of damage and however, in order to use those images with supervised learning technology, each image need to be labeled with the level of damage. The same is being assigned as no damage (0 no sickness) and (1 with sickness).

**2.1 Dataset**

It includes pictures of the lips and tongue that are divided into malignant and non-cancerous categories. Photographs were taken in several ENT facilities in Ahmedabad, India, and were then classified with the aid of ENT specialists. The total cancerous images are 87 while the none cancerous images are 44, full dataset is available on [16].

**2.2 Preprocessing**

Following data processing, each image's records are examined in an effort to determine when the sickness will manifest itself. In this regard, the complete procedure can be summed up as follows:

1) The outcome is a class of "one" and a class of "zero," where "zero" denotes disease or the absence of a fault. Due to the organization of the photo data, each photo has an own class name (which indicate percent sickness occurrence) [5].

2) To reduce training times and lighten the load on the classifiers, image size is being reduced. Scaling an image is necessary for accuracy and to get rid of a lot of superfluous details. This procedure, which consists several coding phases, is utilized several times by the wavelet. We must resize the image for pre-processing in order to eliminate extraneous elements like background noise while keeping the image's proportions [6].

3) The images are normalized which means co images conversion into binary images (black and white) in order to reduce the data load on the training model. To normalize each image, the highest pixel value for a colored image, which is 225, was divided by each pixel [7]. Intelligent classifiers based on convolutional neural networks have been utilized to provide precise disease prediction [8]. The following sections, for instance, offer illustrations of various model settings.

This proposed model, which may be viewed as a more advanced version of an artificial neural network, slightly modified the same recurrent neural network design [9, 10]. It is capable of managing many data values at once. The numerous feedback loops between its layers set it apart from the traditional feed-forward neural network paradigm. A typical (classical) feed-forward neural network's input and output gates are more likely to resemble the input and output layers. On the other hand, the forget layer is the chosen hidden layer in conventional feed-forward neural networks.

**2.3 Features Extraction**

Because image processing applications can automatically and without human intervention extract information from photos, their value has increased [11]. It also has the capacity to reveal information that is hidden from the naked eye. Thus, sophisticated computer programmes and algorithms must be used to process photographs in order to do the necessary duty. This could take several days of training time and a lot of computing power if the image data is huge [12]. To reduce the processing load, techniques like increasing the processor's memory or compressing the photographs by deleting unnecessary data are used. Yet, techniques like wavelet, which are also employed in this paper, can be applied in this situation. Nevertheless, picture coding can be utilized to improve training efficiency while reducing processing burden to the absolute minimum. This method can be used for supervised training if picture target information is given [13].

If electroluminescent image is present with prior target knowledge, such as 30% damage. The goal is to create a comparable, condensed version of this image that has all the necessary information for its identification , makes it stand out from other photographs, and is comparable to the original. In light of this, it is possible to state the following:

== (1)

As demonstrated in (1), is representing the electroluminescent image with id=1 and size of . Which is converted into another matrix of other size (may be varying as per the need). The first row of the new matrix is representing the target information which is in our case reflecting the level of damage (e.g. 30%). The first column is reflecting the image identity e.g. image number 1. In order to mitigate the load on the classifier, the target and identity values are made in binary format. The size of new image was set to . The same is represented in Figure 1.

|  |  |
| --- | --- |
| Actual Image with 30% Defects | 1 0 0 1 0 1 1 0 1 1  1 1 1 1 1 1 1 1 1 1  0 1 1 1 1 1 1 1 1 1  1 1 1 1 1 1 1 1 1 1  0 1 1 1 1 1 1 1 1 1  1 1 1 1 1 1 1 1 1 1  1 1 1 1 1 1 1 1 1 1  0 1 1 1 1 1 1 1 1 1  0 1 1 1 1 1 1 1 1 1  1 1 1 1 1 1 1 1 1 1  Coded Binary Image |

Figure 1: Features codeing demonstration.

**2.4 Model Classification**

To achieve the health diagnosis by tongue image categorization, a model is created using a deep learning approach and is inspired by feed forward neural network and conventional neural network. As a result, a model is developed that forecasts case outcomes using picture analysis in order to establish diagnoses. The model is initially trained with a large image library that comprises a significant number of colored images in order to give the neural network complete knowledge about the object and enable it to handle a range of photos.

Each layer of the model must be trained by allowing it to become popular with images in order for the network to use the knowledge acquired from all of the layers to recognize an image. As a result, the model may correctly predict the diagnosis during the test phase.

**2.4.1 Model Training**

The model will receive additional instruction until it is performing at its peak. In order to keep the error in the results from increasing, the optimization strategy may recalculate the weight until it achieves the lowest error in the output. The integrated training algorithm may evaluate the precision of the outcomes to gauge the efficacy of the neural network model [14].

The evaluation of the model's performance using the fitness function enables the determination of the final weight coefficients [15]. The initial stages of training involve producing random weight coefficients and assigning those figures to the weight values of the neural network. The training process is therefore continued throughout the training phases in order to achieve weight values that minimize the fitness function. In this case, the fitness function is represented by the mean square error. To begin testing, test data are provided to the neural network model's input, which seeks to ascertain the state of health. The neural network will assess the input data before making a prediction about the image class. This model was created to produce the diagnosis report and incorporate the dataset's photos. Figure 2 depicts the whole process proposed in this work.

1. **RESULTS**

According to the obtained results from the above methods which is summarized in Table 1, the following points can be made.

(a) That maximum accuracy is 100 percent which is achieved in both proposed classifiers e.g. FFNN and CNN when Coding technique is used in Features extraction model.

(b) The minimum accuracy is seen in FFNN model when wavelet transform is used in features extraction.

(c) In case of wavelet features extraction, CNN is outperformed over the FFNN model in terms of time (248.232 percent) is achieved. However, the maximum accuracy is achieved in CNN in account of time where time required in CNN is 248.232 seconds for achieving of two-fold of FFNN accuracy.

(d) Coding techniques in features extraction is optimized both classifiers accuracy to 100 percent. Time required to perform the classification in the mentioned accuracy in FFNN is far less than it in case of CNN. Consequently, single hidden layer FFNN is outperformed when coding technique is used by obtaining of 100 percent accuracy of classification at 0.95993 seconds. FFNN is leading the other classifier in both accuracy of classification and time.

Table 1: performance assessment.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tools | AC measure | MSE measure | Time measure | MSE measure |  |
| FFNN | WL | 33.333 | 1.8410 | 62.3201 | 1.00387 |
|  | Coding | 100 | 0 | 0.95993 | 0 |
| CNN | WL | 63.953 | 0.3604 | 248.232 | 0.36046 |
|  | Coding | 100 | 0 | 208.236 | 0 |

The same is graphically represented at Figure 3 .



Figure 3: accuracy measure of the tongue image classifier.

1. **CONCLUSION**

There have been various attempts to employ texture tracking features to diagnose diseases including eczema, scarlet fever, and the fairly common Kawasaki disease, which causes changes in tongue colour. All alleged detection techniques have been demonstrated to have a number of flaws, such as (time delay and high computational cost). The technology examines the body using photos of the tongue to discover disorders including Kawasaki disease, eczema, and scarlet fever that are invisible to the human eye. According to this study, automated anomaly detection would be a considerably more cost-effective alternative to human monitoring for carrying out the task of detection and ensuring excellent performance. Performance of the suggested state-of-the-art is compared to that of the Convolutional Neural Network (CNN) and Feed Forward Neural Network, for example, the Feed Forward Neural Network (FFNN). The proposed state of the art, such as coding-based features extraction, performs better in terms of abnormality/defect identification. The single hidden layer FFNN was able to attain the maximum level of recognition accuracy of 100% in the lowest amount of time, 0.95993 seconds, whereas both classifiers were capable of doing so. When the performance of the recommended model was assessed using tenfold validation, the aforementioned accuracy is the highest one that was attained in the fifth fold. The proposed state-of-the-art technique yields findings with the lowest mean absolute error (MAE), which is assessed for both techniques (zero).

**REFERENCES**

[1] T.J.W. Klein Nulent et al. Intraoral ultrasonography to measure tumor thickness of oral cancer: a systematic review and meta-analysis, Oral Oncol (2018)

[2] Y. Shinozaki et al. Relationship between appearance of tongue carcinoma on intraoral ultrasonography and histopathologic findings, Oral Surg Oral Med Oral Pathol Oral Radiol (2014)

[3] A. Yesuratnam et al., Preoperative evaluation of oral tongue squamous cell carcinoma with intraoral ultrasound and magnetic resonance imaging—comparison with histopathological tumour thickness and accuracy in guiding patient management, Int J Oral Maxillofac Surg (2014)

[4] A. Kaneoya et al., Quantitative analysis of invasive front in tongue cancer using ultrasonography, J Oral Maxillofac Surg. (2009)

[5] S. Shintani et al., Intraoral ultrasonography is useful to evaluate tumor thickness in tongue carcinoma. Am J Surg. (1997)

[6] T.M. Pimenta Amaral et al., Predictive factors of occult metastasis and prognosis of clinical stages I and II squamous cell carcinoma of the tongue and floor of the mouth. Oral Oncol. (2004)

[7] L. Lam et al., Retrospective study of survival and treatment pattern in a cohort of patients with oral and oropharyngeal tongue cancers from 1987 to 2004. Oral Oncol (2007)

[8] T. T. Deng, “Basic theory of traditional Chinese medicine,” in Diagnostics of Chinese Medicine, pp. 5–11, Chih-Yin Publishing, Taipei, Taiwan, 1995.

[9] C. H. Li and P. C. Yuen, “Tongue image matching using color content,” Pattern Recognition, vol. 35, no. 2, pp. 407–419, 2002.

[10] Y. Wang, Y. Zhou, and J. Yang, “A tongue analysis system for tongue diagnosis in traditional Chinese medicine,” in Proceedings of the International Symposium Computational and Information Science (CIS ’04), vol. 3314 of Lecture Notes in Computer Science, pp. 1181–1186, Springer, 2004.

[11] 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, vol. 28, no. 1, pp. 11–19, 2007.

[12] Q. Li and Z. Liu, “Tongue color analysis and discrimination based on hyperspectral images,” Computerized Medical Imaging and Graphics, vol. 33, no. 3, pp. 217–221, 2009.

[13] M. F. Zhu and J. Q. Du, “A novel approach for color tongue image extraction based on random walk algorithm,” Applied Mechanics and Materials, vol. 462-463, pp. 338–342, 2014.

[14] B. Pang, D. Zhang, and K. Wang, “Tongue image analysis for appendicitis diagnosis,” Information Sciences, vol. 175, no. 3, pp. 160–176, 2005.

[15] C.-C. Chiu, “A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue,” Computer Methods and Programs in Biomedicine, vol. 61, no. 2, pp. 77–89, 2000.

[16] Dataset: https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images