

Mononucleosis oral disease detection with the help of GANNs

Mihai Gherghinescu, Todor Ivaşcu^[0000–0003–3405–7358], and Sebastian Ştefăniş^[0000–0002–6211–9205]

Department of Computer Science, West University of Timisoara
Blvd. V. Pârvan 4, Timișoara 300223, Romania
{mihai.gherghinescu00,todor.ivascu,sebastian.stefaniga}@e-uvvt.ro
<https://www.info.uvt.ro>

Abstract. This paper aims to provide a new approach of automatically classifying common throat diseases, with a particular emphasis for mononucleosis. An initial exploration will focus on the impact of artificial intelligence (AI) on oral disease classification. Classification methods that have proven to be successful will be presented, along with a description of how processing both domain-specific and general oral cavity images can assist doctors in automatically assessing patient prognosis. We will begin by presenting a selection of papers that we believe reflect the current state of the art in this field of automatically detecting oral diseases. These studies will provide a foundational understanding of the existing approaches and techniques. The proposed methods will be analyzed in greater detail, with a particular emphasis on evaluating the robustness of the training data and the training procedure, as these factors are crucial for ensuring the accuracy and reliability of the models. During our approach we will highlight the lack of good publicly available datasets and the need for a new dataset for our particular problem. The data collection and processing procedure will be further described as well as the data augmentation process. The final dataset contains 832 images that have been labeled with the help of Students from Victor Babes University of Medicine and Pharmacy Timisoara. Data is split in four classes: healthy, tonsillitis, pharyngitis and mononucleosis, three of them representing oral diseases. It contains augmented data, 382 images were generated with the help of Generative Adversarial Neural Network. Furthermore, we will present the process of designing and training our model while using the collected data. Final model is represented by an ensemble of three models and achieved 93.0% accuracy and 93.2% recall. In the end, we will summarize our results, providing a comparison to the identified current state of the art. This will include a critical evaluation of the limitations of our approach. Furthermore, we will present potential improvements aimed at boosting model accuracy as well as discuss the possibility of introducing new classes to enhance the model's ability to classify a wider range of oral diseases.

Keywords: Oral disease, Convolutional Neural Networks, Mononucleosis, Data Augmentation, Generative Adversarial Neural Network

1 Introduction

Traditional diagnostic approaches often pose challenges in terms of accessibility, cost, and time, hindering timely intervention and exacerbating the impact of oral health issues. The introduction of AI into the field of oral health not only addresses these shortcomings but also opens up new avenues for enhanced precision and objectivity in diagnostics. By leveraging machine learning and pattern recognition algorithms, AI has the potential to revolutionize the accuracy and applicability of oral disease detection, providing a valuable tool for healthcare professionals, especially when dealing with large amount of data.

One specific example is that during the recent COVID-19 pandemic the hospitals had to deal with a large number of patients as presented in [1] leading to delayed treatments for patients. There was a high strain on doctors due to the large volume of data that needed to be processed and they had to prioritize the patients having COVID-19. There are also cases when people were hesitant of visit the hospital facilities due to the fear of contamination. To solve this issues there needed to be an reliable alternative way that the ill could self diagnose from the safety of their homes. This way, only relevant cases would eventually arrive to the hospital facilities and be diagnosed, reducing the doctors volume of work. Some life changes that have been made are presented in [4].

Another important factor is that while medical professionals strive for accuracy in diagnoses, the complexity of human health can lead to occasional misdiagnoses, even among the most skilled practitioners. Recognizing this inherent challenge, the integration of Artificial Intelligence in healthcare holds promising potential. AI can serve as a valuable tool, providing a second opinion on patients' health conditions diagnosis. By leveraging advanced algorithms and machine learning, AI systems can analyze medical data with speed and precision. This capability allows them to cross-reference symptoms, medical history, and diagnostic images to generate additional insights. In scenarios where misdiagnosis is a concern, AI acts as a safeguard, offering an independent evaluation of the patient's condition. AI systems are not immune to errors, but they bring an additional layer of consistency to the diagnostic workflow.

To provide a comprehensive overview of the evolving landscape at the intersection of technology and oral health, we will commence our exploration by presenting a curated selection of scientific papers. Our selection process involved a meticulous examination of seminal contributions, ensuring a representative coverage of advancements in detecting dental diseases, oral cancer, pharyngitis, and tonsilo-pharyngitis. The chosen papers not only showcase the progression of research in the field but also highlight the diversity of technologies employed over time. We will delve into each paper, offering insights into the technologies developer, their respective applications and focus on the quality of data used.

Next, we will highlight the lack of good publicly available datasets and the need for a new one, tinkered for our particular problem. The data collection and processing procedure will be further described as well as the data augmentation process. We will delve into the concept using using GANNs and present the model architecture that has been trained to generate more images.

Based on this dataset, containing only publicly available data, we will propose an ensemble composed of three sub-models used for detecting 3 diseases: tonsillitis, pharyngitis and mononucleosis. We will present the training procedure and argument why the class distribution is balanced. We will highlight and describe the performance metrics of the ensemble, as well as visualize predictions. Final metrics achieved are 93.0% accuracy and 93.2% recall.

In the end, we will summarize our results, providing a comprehensive comparison to the current state of the art in the field. This will include a critical evaluation of the strengths and limitations of our approach relative to existing methods. Furthermore, we will present potential improvements aimed at boosting model accuracy, such as optimizing model architectures, or incorporating additional features. Additionally, we will discuss the possibility of introducing new classes to enhance the model’s ability to classify a wider range of oral diseases, thereby broadening its applicability and improving its clinical utility.

2 Related Work

Figure 1 shows that the popularity of the top four most researched fields within the domain of AI-enhanced oral health disease diagnosis. We can see that the majority of papers are focused on dental disease detection or oral cancer detection as both conditions have significant health implications. Additionally, we observe a steady interest in detecting tonsillitis and pharyngitis, indicating a broader focus on addressing common throat-related conditions using AI technologies.

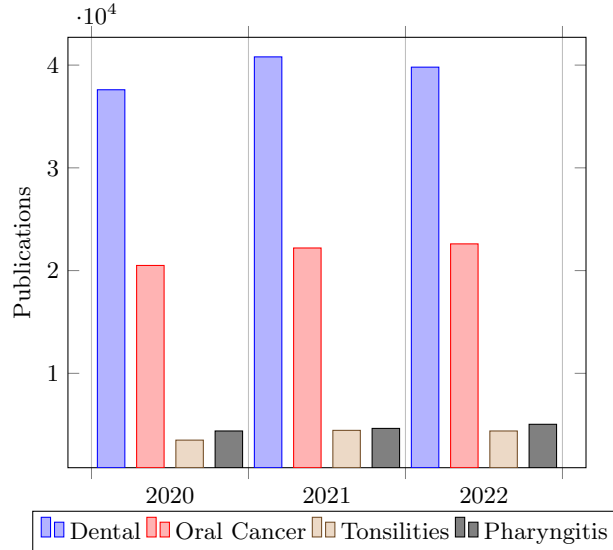


Fig. 1: Popularity of Domains based on number of publications

In the following section we will present a selection of papers that we believe represent the current state of the art. The selection criteria was based on the number of citation, how recent the paper is and the conference rankings.

2.1 Dental Diseases Detection

Classification of dental diseases using CNN and transfer learning [5]

The paper focuses on addressing multiple dental issues, including dental caries, periodontitis, and periapical infection. The research utilizes a private dataset comprising of 251 x-ray RVG images. For the analysis, the authors employ the use of a Convolutional Neural Network architecture: VGG16. The authors claim that the final model accuracy is 88.46%, but do not seem to specify the number of runs used to evaluate the model accuracy.

Our main concerns with this paper revolve around the dataset utilized in the study. Specifically, the authors noted that the data distribution consists of 80 dental caries images, 110 periapical infection images, and 61 periodontitis images. This distribution indicates a significant class imbalance, with the periodontitis category containing nearly half the number of samples compared to the periapical infection category. Such imbalance can negatively impact the model performance as well alter the true evaluation metrics. Also, the dataset size is quite small and insufficient to confidently acknowledge the training results, accuracy being calculated only over 26 images.

Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm [3] Compared to the previous paper, this paper focuses solely on dental caries, but uses a much larger private dataset composed of 3000 periapical radiographs that is further augmented using basic data augmentation techniques. The dataset seems to be split in 4 balanced classes resembling the type of tooth being diagnosed and it's affection: healthy premolar, premolar caries, healthy molar, molar caries. As for the model architecture they used InceptionV3 and achieved 89.0% accuracy and 84.0% recall.

As for future improvements the authors presented some flaws like the resolution limitations of the photos being used. Another limitation would be the number of samples used as well as the type of teeth captured by the images.

2.2 Oral Cancer Detection

Automatic classification and detection of oral cancer in photographic images using deep learning algorithms [6]

The main focus of the paper is on oral cancer detection. Oral cancer stands as a lethal disease, ranking among the prevalent malignant tumors globally. The research employs a dataset consisting of 700 clinical images for analysis composed of 2 classes: oral squamous cell carcinoma and normal oral mucosa. The method of choice for classifying the data is to use a Recursive Convolution Neural Networks based on DenseNet121 architecture. The authors claim that the model has achieved precision of 99%,

a recall of 100%, and an F1 score of 99%. The method of choice to measure the performance is K-fold cross validation with a fold of five with an 80/20 split.

The only concern with this paper is the almost perfect accuracy that might be due to over-fitting raised from the fact that the authors did not specify the number of training steps or provide a time-chart presenting the training process.

Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm [2]

Like the paper presented beforehand, this paper aims to provide a way to classify oral cancer. As for the training data the authors combined 3 different dataset, collecting a total of 1300 images. As for the model being used they chose InceptionV3 architecture and achieved 94.5% accuracy.

The main concern is related to the training process. First of all the authors didn't specify the data distribution split as well as the number of epochs or test repetitions. Second of all the dataset is truly imbalanced, there are 1140 tumor images and 160 healthy images. The dataset is composed of 87% unhealthy images and 160 healthy images leaving us to believe the the training results are invalid as the model choosing tumor all the time leads to 87% overall accuracy.

Also, there are many unjustified statements like "the challenges to designing an accurate classifier are: It is able to process a real time series data in available Random-Access Memory (RAM)". Why the problem of local RAM would apose a problem is still unknown and we believe it could of been easily solved in most cases.

2.3 Pharyngitis Detection

Toward automated severe pharyngitis detection with smartphone camera using deep learning networks [8]

The main focus of the paper is on pharyngitis automatic detection. The research utilizes a dataset comprising of 339 Google images. For the analysis, the authors employ the ResNet50 neural network architecture. The results indicate an accuracy of 95.3%. Due to lack of data authors tried to train a GANN to generate more images. They used an unconventional approach, GANN having as input the healthy/unhealthy images and producing the opposite class.

There are two big concerns related both to the training data as well ass the training procedure. First of all the authors specified that they used data augmentation techniques to reach a total of 3885 images, but the data distribution during the training procedure seems to be imbalanced. To be more precised, in the validation set they used 500 healthy images and 300 unhealthy images, the number of healthy images being 1.6 times hight then the unhealthy ones, leading to great class unbalances that denaturate the final model performance metrics.

Another big concern is the fact that during the training, even though k-fold validation with a fold of 4 was used for testing, the data distribution was 95/5, leaving only 85 images for testing out of the 3885. Given the fact that the initial dataset was basically multiplied with 10 replicas for each image, there is a very

high probability the training and testing data contain the same samples. This would lead to the model actually learning the dataset and overfitt.

The last concern is that the authors didn't mention the help of any professionals when it comes to data labeling. This raises a big question regarding the actual validity of the training dataset. Even the augmented samples generated by the GANN do not seem to actually resemble pharyngitis, but rather another disease, mononucleosis, due to the presence of white dots.

2.4 Tonsillopharyngitis Detection

Detection of Tonsillopharyngitis with Grad-Cam and Optimization-Based Model Using Oropharyngeal Images [7] Like the previous paper, the authors seem to used the exact same dataset composed of 339 images, but rather detect tonsillopharyngitis. Rather then focusing on one model, the model tries to compare different architectures: GoogleNet, AlexNet, DenseNet201 and DarkNet53. As for the final model being used, they chose Darknet53 architecture and made use of Grad-Cam images. It managed to achieve 89.7% accuracy.

The main concern is related to the data class imbalance. The dataset consists of 131 throats and 208 normal throat images, leading to great class imbalance and denaturated model performance results. Also the capability of the model being able to reliably generalize is questionable due to the small dataset size.

3 Open problems

3.1 Mononucleosis

The topic that we will continue with within this paper, that has yet to be discussed in any papers, is providing an automatic mononucleosis diagnostic system using artificial intelligence. Mononucleosis, also known as the "kissing disease," is a viral infection primarily caused by the Epstein-Barr virus (EBV). The infection is characterized by a group of symptoms that can be mistaken for other illnesses, making accurate and timely diagnosis essential.

Symptoms of mononucleosis often include extreme fatigue, fever, sore throat, swollen lymph nodes, and an enlarged spleen. In some cases, liver inflammation (hepatitis) and jaundice may also occur. The symptoms can vary in intensity and duration, with some patients recovering in a few weeks and others taking several months to fully regain their energy.

Diagnosis of mononucleosis typically involves imaging studies but can further lead to a combination of clinical assessment, laboratory tests (such as a complete blood count to detect atypical lymphocytes and heterophile antibody tests like the Monospot test) if complications are suspected.

4 Proposed approach

4.1 Data collection and classification

Due to the lack of good publicly available datasets we took the challenge of building our own. We started by collecting data from publicly available datasets from Roboflow, Kaggle, Google Images and Bing Images. In this process we mostly searched for images that contained the affected areas of the oral cavity, not necessarily on a particular disease. After scrapping everything we could find, we ended up with a preliminary dataset of 3000+ images that contained a lot of noise, badly placed images and was not labeled.

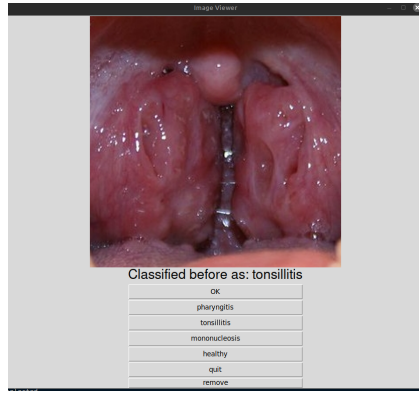


Fig. 2: Helper tool for data classification

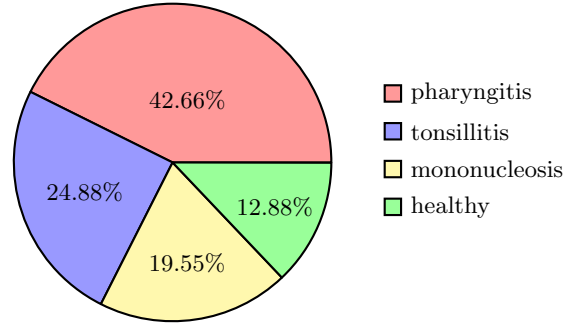


Fig. 3: Class distribution before data augmentation

After collecting the preliminary data set, we went on several iterations of classifying and curating the data set. There are four classes that the dataset is split into: healthy, tonsillitis, pharyngitis, and mononucleosis. Due to the need for professional expertise in classifying diseases, the labeling itself was done with the help of students from Victor Babes University of Medicine and Pharmacy Timisoara by using a helper app built to ease their work 2. Additionally, we did some further preprocessing on the images to boost model performance. Images are cropped to contain only the affected region of the oral cavity due to reduce the volume of noise that was greatly affecting our training results.

After finishing curating and labeling the dataset we were left with 450 images. To our surprise we managed to find more unhealthy individual images than healthy ones, to be more precise figure 3 highlights the data distribution of the dataset before performing data augmentation. As can be seen from the chart, the least amount of samples, 58 images, were represented by the "healthy" class.

4.2 Data augmentation and an short overview of Generative Adversarial Networks

To be able to bypass the limitations of having a small amount of healthy individual images we tried using data augmentation techniques to artificially generate more images. We ended up building a Generative Adversarial Neural Network that was capable of generating artificial samples.

Generative Adversarial Neural Networks (GANs) are a class of machine learning models introduced by Ian Goodfellow in 2014. GANs consist of two neural networks: the **generator** and the **discriminator**, which work in opposition to each other. The generator's role is to create fake data (such as images) that resemble real data, while the discriminator's task is to distinguish between real and generated (fake) data.

The **generator** takes random noise as input and tries to produce data that mimic the distribution of real data, such as generating images of faces that look like real human faces. The **discriminator**, on the other hand, is trained to differentiate between real and fake data by being shown both real examples and the data generated by the generator. The two networks are trained simultaneously in a *game-theoretic* manner:

- The generator improves the produced data to fool the discriminator.
- The discriminator improves by learning to better identify the fake data.

This adversarial process continues until the generator produces data that is nearly indistinguishable from real data, and the discriminator cannot reliably distinguish between the two. The ultimate goal of GANs is to generate realistic data that can be used in various applications, such as image generation, video creation, and even data augmentation.

4.3 GANN architecture

To further enhance our dataset, we build a GANN that is capable of generating 160*160 RGB healthy images. For training we used the provided 58 healthy images presented in our initial dataset and stopped when the discriminator loss was close to 0 and the generator couldn't keep up anymore, it's loss oscillating between 1.2 and 5. We didn't use any predefined models but rather build our own architecture from scratch that would fit our problem. We ended up having in appearance a much more powerful generator, containing 10x the amount of parameters the discriminator has. As for the results achieved pictures 4 and 5 illustrates some of the generated images.

Figure 7 represent the discriminator architecture and figure 6 represents the generator architecture. The generator is composed of seven repetitive Conv2DTranspose and Normalization for up-sampling our data. The discriminator is doing the opposite, down-scales the data using three Conv2D layers, flattens the values and has as output the confidence score.

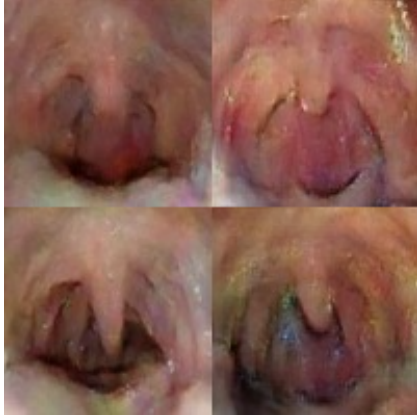


Fig. 4: Generated images samples



Fig. 5: Generated images samples

4.4 Model architecture and class distribution

The final dataset 8 contains 832 images, from which 382 were synthetically generated with the help of the proposed GAN architecture. Despite the augmentation process, the dataset remains unbalanced, which poses challenges for the model's ability to effectively classify underrepresented classes. Given the difficulty in acquiring more real data samples, we decided to implement an ensemble model composed of three separate models to address the issue of class imbalance. Each of these models functions as a binary classifier, and the ensemble as a whole operates in a chain-of-thought fashion, with each stage focusing on progressively narrowing down the possible diagnoses.

The classification approach is designed to gradually determine the condition of the patient. This modular approach allows for better handling of the imbalanced data by focusing on smaller, more manageable subsets of the overall problem. In terms of architecture, we experimented with several well-known deep learning models for each of the sub-models, including InceptionV3, VGG16, ResNet (V2), and EfficientNet. These models were selected due to their proven performance in image classification. In the next section we will highlighting only the best performing model architectures.

The first model detects two classes "unhealthy" and "healthy" and classifies all 832 images within the dataset that are not labeled as "healthy" as "unhealthy". The data distribution can be seen in figure 9.

The second one tries to determine if the affected area is scoped to the tonsils or to the pharynx and uses 382 images from the dataset. We classify "mononucleosis" as a tonsils related issue, the data distribution can be seen in figure 10.

The last one is applied only if a tonsil related issue was found and classifies the disease in 2 categories: mononucleosis and tonsillitis. As for the training data we used 238 images, represented by the already labeled data, the data distribution can be seen in fig. 11

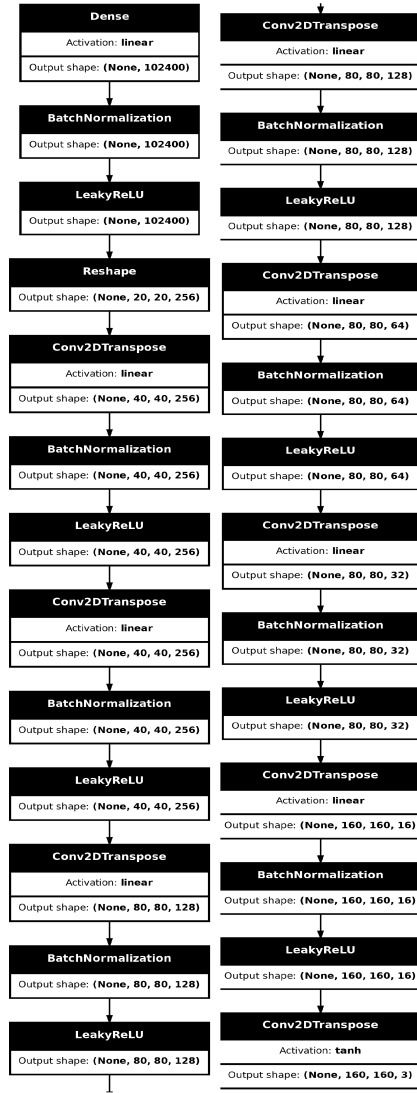


Fig. 6: Generator model architecture

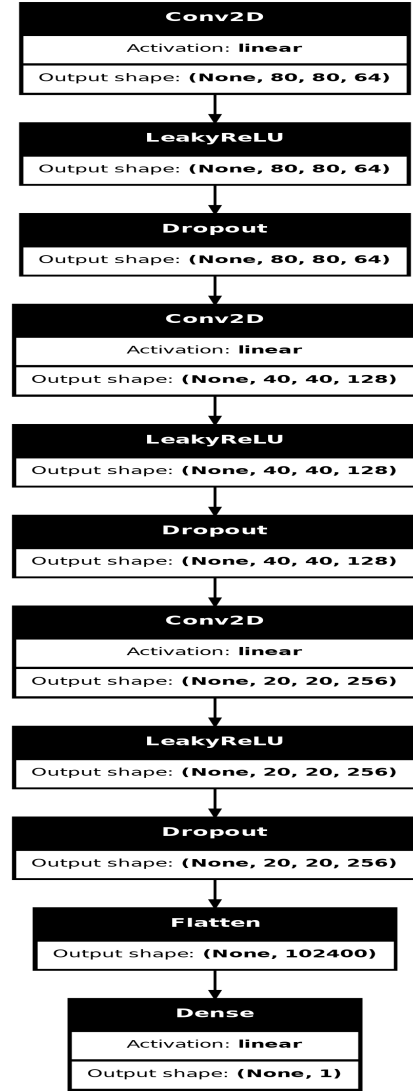


Fig. 7: Discriminator model architecture

5 Results

The final ensemble model achieved accuracy 90%, 91% precision, 90% recall, 93% roc-auc, 90% f1-score. As for the training, we randomly did an 80-10-10 split and made sure data was distributed evenly in concordance to the class it belonged to. After that we applied K-Fold validation to test our metrics, with a fold of 5, leaving 10% of the data for testing. We repeated the process several more times

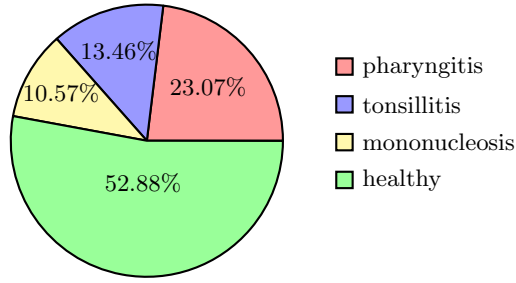


Fig. 8: Class distribution after data augmentation

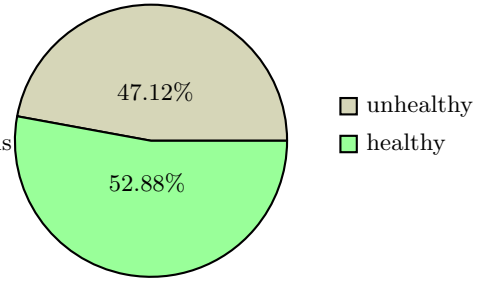


Fig. 9: Data distribution for the first model

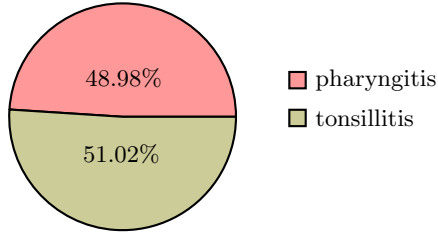


Fig. 10: Data distribution for the second model

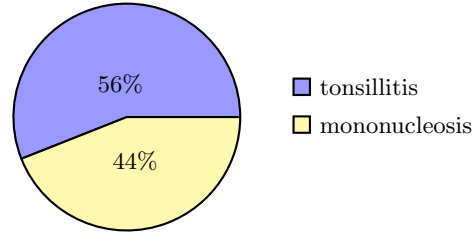


Fig. 11: Data distribution for the last model

and took the average as the final result. For the model architecture, we ended up using a common one for all the models after testing several possibilities. Next, we will describe the architecture used and highlight the process of training the models and display the performance results. As a note, the faded lines represent the un-smoothed values that form the time series in the following charts.

The base architecture used is based on a predefined InceptionV3 model and uses transfer learning. As for the input layers, we first have a Resizing layer to upscale our data to size (299, 299, 3) due to the amount of layers of convolution the base architecture has. We then froze all the models pre-trained layers. We added one GlobalAveragePooling2D layer to make the transition to the prediction layer and added one Dense, 128 nodes, layer using leaky relu with and alpha of 0.1. During our training we also added an EarlyStopping callback as well as ReduceLROnPlateau callback to prevent over-fitting.

The 'healthy-unhealthy' model achieved 93% accuracy, 94% precision, 93% recall, 94% roc-auc, 93% f1-score. Figure 12a represents the accuracy increase during the training process based on the epoch. The blue line is the validation accuracy and the black line represents the training accuracy.

The 'Pharyngitis-Tonsillitis' model achieved 95% accuracy, 95% precision, 95% recall, 93% roc-auc, 95% f1-score. Figure 12b represents the accuracy in-

crease during the training process based on the epoch. The pink line is the validation accuracy and the blue line represents the training accuracy.

The 'Tonsillitis-Mononucleosis' model achieved 89% accuracy, 90% precision, 88% recall, 89% roc-auc, 88% f1-score. Figure 12c represents the accuracy increase during the training process based on the epoch. The orange line is the validation accuracy and the black line represents the training accuracy.

Table 1 presents a comparison between the previously presented papers and ours, highlighting the model architecture, methods used for the classification task and the data that has been feed to the model as well as it's performance. The proposed model surpasses 50% of the presented papers in terms of accuracy and has more data to train on. Also, in contrast to all the other papers, our paper clearly states and argues that the dataset is balanced as well as provides f1-score as well as roc-auc-score to highlight the fact that the classes are balanced.

6 Conclusions and Future Work

Ref.	Addressed Problem	Data Used	Methods Used	Results
[5]	Dental caries, periodontitis, periapical infection	251 X-ray RVG images	VGG16 (CNN)	88.46% accuracy
[3]	Dental caries for premolars and molars	3000 periapical radiographs	GoogLeNet Inception v3 (CNN)	86.0% accuracy
[6]	Oral cancer detection	700 clinical images	DenseNet121, R-CNN	100% accuracy
[2]	Oral cancer detection	1300 hyperspectral images	Augmented GoogLeNet Inception v3 (CNN)	94.5% accuracy
[8]	Pharyngitis	339 Google images + 3546 generated images	ResNet50	95.3% accuracy
[7]	Tonsillopharyngitis	339 Google images	Grad-Cam and Darknet53	89.7% accuracy
Ours	Pharyngitis, tonsillitis, mononucleosis	450 public images + 382 generated images	Ensemble: InceptionV3, InceptionV3, VGG16	90.0% accuracy

Table 1: Comparison between current state-of-the-art methods and our results.

This paper presented a novel approach to the detection of mononucleosis by leveraging the combined strengths of Generative Adversarial Neural Networks (GANs) and Convolutional Neural Networks (CNNs). The methodology focused on providing a new model architecture that could cover the unresearched task of automatically detecting mononucleosis through the help of image classification. The results showed that, given the current small and dissipated available data, we managed to collect enough data samples to train our model. Also, data augmentation techniques like GANs proved to be useful when attempting to further enhance the dataset size and also balance out our dataset. We managed to provide a curated dataset containing 832 images, from which 382 were artificially generated with the help of GANs. The labeling was done with the help

of medicine university students providing further credibility to our work. The dataset surpasses in size and data quality any other publicly available datasets currently available and will be, in the future, intended for public usage.

As for the model architecture being used we provided a better way to split the detection process, that improved overall model accuracy, by combining multiple binary classification models in a multi-class classification task. This also allowed us to equally split the data despite the fact that the number of images for each class is different.

Despite these promising results, there are several areas where further work could enhance the model and extend its applicability. First, the current study focused on a relatively small dataset of medical images, which, although augmented with GANNs, may still limit the generalizability of the model to larger, more diverse populations. Future research should aim to validate the model on larger, multi-institutional datasets to ensure broader applicability and robustness across different demographic and geographic regions. Additionally, while the use of GANNs for data augmentation proved beneficial, exploring more advanced generative models, such as Variational Autoencoders (VAEs) or improved versions of GAN architectures, could lead to even more realistic synthetic data generation and further improvements in model performance.

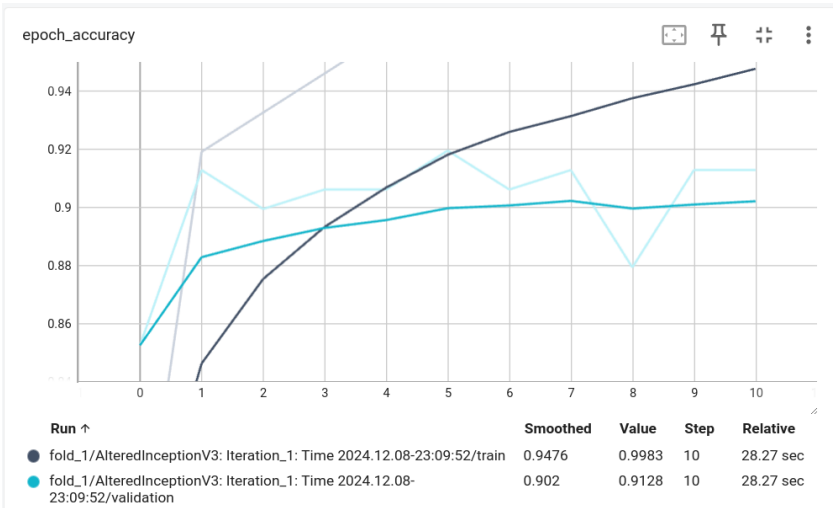
Another potential avenue for future work lies in the refinement of the CNN architecture. While the current model architecture performed well, during the training process we observed that the model is slightly over-fitting on the training data, especially due to the number of training samples.

Lastly, the application of this method could be extended beyond mononucleosis detection to other medical conditions. The framework proposed in this paper can serve as a foundation for developing advanced detection systems for a wide range of diseases, particularly in fields where high-quality labeled datasets are scarce. Future studies could explore this framework’s adaptability to other domains in medical imaging and its potential for real-world deployment.

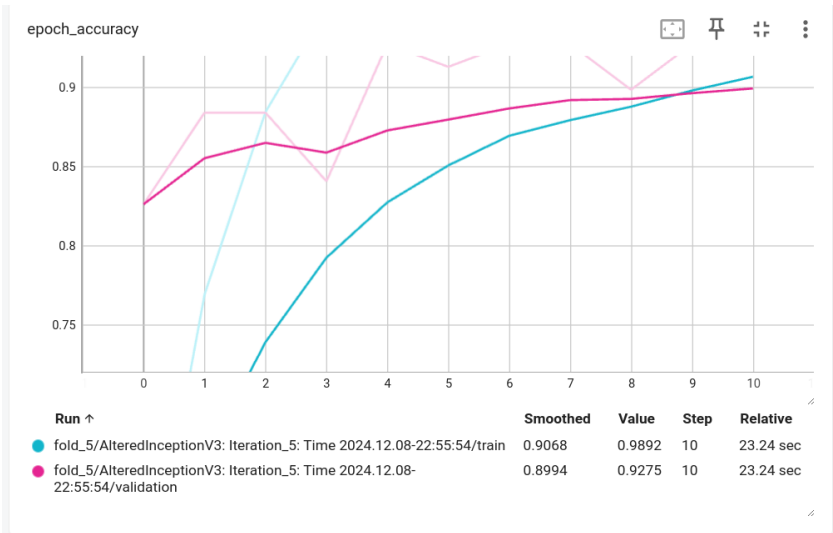
In conclusion, this research demonstrated the effectiveness of combining GANNs and CNNs for mononucleosis detection, achieving promising results. The synergy between the two models allowed for enhanced feature extraction and classification, showcasing their potential to address complex medical diagnostic tasks. However, despite these promising outcomes, there are still challenges to overcome. Continued efforts in data expansion through the inclusion of diverse and larger datasets are essential to improve generalizability and robustness. Additionally, further refinement of the model architecture and hyperparameters could lead to improved accuracy and efficiency. Finally, broader validation across different patient populations and clinical scenarios is crucial to ensure reliability and applicability in real-world medical settings. This study serves as a strong foundation for future research, opening pathways for innovative AI-driven approaches in healthcare diagnostics.

References

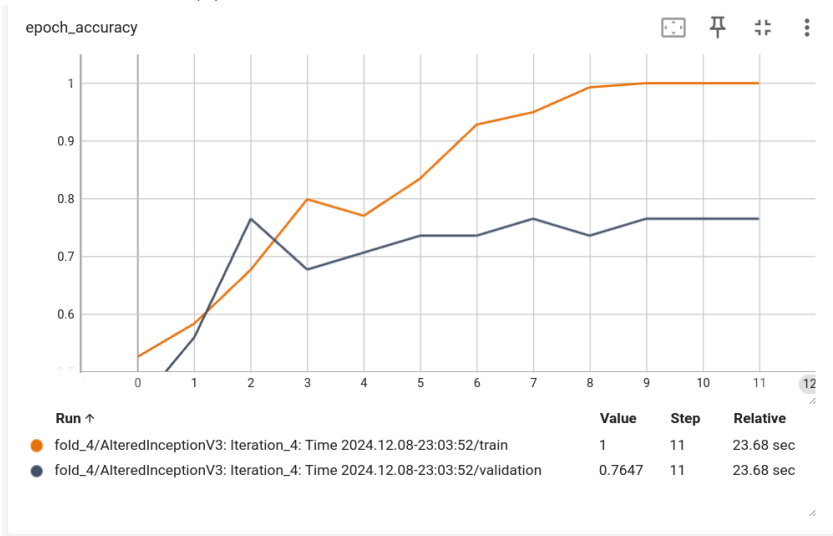
1. French, G., Hulse, M., Nguyen, D., Sobotka, K., Webster, K., Corman, J., Aboagye-Nyame, B., Dion, M., Johnson, M., Zalinger, B., Ewing, M.: Impact of hospital strain on excess deaths during the covid-19 pandemic — united states, july 2020–july 2021. *MMWR. Morbidity and Mortality Weekly Report* **70**(46), 1613–1616 (Nov 2021). <https://doi.org/10.15585/mmwr.mm7046a5>
2. Jeyaraj, P.R., Samuel Nadar, E.R.: Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *Journal of Cancer Research and Clinical Oncology* **145**(4), 829–837 (Jan 2019). <https://doi.org/10.1007/s00432-018-02834-7>
3. Lee, J.H., Kim, D.H., Jeong, S.N., Choi, S.H.: Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of Dentistry* **77**, 106–111 (Oct 2018). <https://doi.org/10.1016/j.jdent.2018.07.015>
4. Maison, D., Jaworska, D., Adamczyk, D., Affeltowicz, D.: The challenges arising from the covid-19 pandemic and the way people deal with them. a qualitative longitudinal study. *PLOS ONE* **16**(10), e0258133 (Oct 2021). <https://doi.org/10.1371/journal.pone.0258133>
5. Prajapati, S.A., Nagaraj, R., Mitra, S.: Classification of dental diseases using cnn and transfer learning. In: 2017 5th International Symposium on Computational and Business Intelligence (ISCBI). IEEE (Aug 2017). <https://doi.org/10.1109/iscbi.2017.8053547>
6. Warin, K., Limprasert, W., Suebnukarn, S., Jinaporntham, S., Jantana, P.: Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *Journal of Oral Pathology & Medicine* **50**(9), 911–918 (Aug 2021). <https://doi.org/10.1111/jop.13227>
7. YILDIRIM, M., EROĞLU, O.: Detection of tonsillopharyngitis with grad-cam and optimization-based model using oropharyngeal images. *NATURENGS MTU Journal of Engineering and Natural Sciences Malatya Turgut Ozal University* (Dec 2022). <https://doi.org/10.46572/naturengs.1205919>
8. Yoo, T.K., Choi, J.Y., Jang, Y., Oh, E., Ryu, I.H.: Toward automated severe pharyngitis detection with smartphone camera using deep learning networks. *Computers in Biology and Medicine* **125**, 103980 (Oct 2020). <https://doi.org/10.1016/j.compbio.2020.103980>



(a) Healthy-Unhealthy model accuracy



(b) Pharyngitis-Tonsillitis model accuracy



(c) Tonsillitis-Mononucleosis model accuracy

Fig. 12: Model accuracies for different classifications.