Towards Graph-based Semi-supervised Learning on Audio Embeddings for Label Classification

Rishabh Chandaliya*, Mohan Timilsina, John Breslin and Martin Serrano
Department of Computer Science - University of Galway, Ireland
Insight SFI Research Centre for Data Analytics
Galway, Ireland * Corresponding Author
[rishabh.chandaliya, mohan.timilsina]@insight-centre.org
and [john.breslin, jamiemartin.serranoorozco]@universityofgalway.ie

***Abstract*—The increasing importance of audio-based healthcare diagnostics, particularly in chronic respiratory problems, has prompted the development of novel approaches for disease classification using the full spectrum analysis of cough sounds. This paper presents an innovative study exploring the potential of employing supervised and semi-supervised learning methodologies for disease categorization based on cough audio samples. Specifically, this study focuses on scenarios characterized by a shortage of annotated data related to chronic diseases. The objectives of our study involved the utilization of standard machine learning algorithms for direct classification based on embeddings, as well as the integration of Graph Neural Networks (GNNs) on the KNN graph. Preliminary results indicate that GNN models consistently outperformed traditional classifiers. For instance, with just 1% of the data, GAT and GCN achieved AUC PR values of 0.84 and 0.87, respectively, surpassing all traditional methods. The superiority mentioned above was sustained even when the data fraction was augmented to 3% and 5%. The usefulness of graph neural networks (GNNs) was further supported by comprehensive performance measures, wherein the graph convolutional network (GCN) exhibited exceptional precision and PR (precision-recall). In summary, the AudioVec-Diagnosis framework presents a promising opportunity for further investigation in audio-based healthcare diagnostics. It provides an optimum approach for situations with a limited availability of labeled data.**

***Index Terms*—Audio-based medical diagnostics, Cough noises, Semi-supervised learning, Disease classification, AudioVec-Diagnosis, Wav2Vec2.0 model**

I. Introduction

The COVID-19 pandemic is still having a profound and far-reaching effect on a global scale. Declared in Q1-March 2019 as global pandemic and by Q3-August 2021, the global tally of confirmed COVID-19 cases surpassed 202 million, with a corresponding death toll exceeding 4 million. In response to the situation, several nations opted to enforce lockdown measures, resulting in economic recessions and challenges to mental well-being, including melancholy and stress symptoms [1] [2]. The importance of having an accurate and timely diagnosis cannot be emphasized more than ever. A frequently utilized diagnostic technique is reverse transcription polymerase chain reaction (RT-PCR). Nevertheless, it is essential to acknowledge that RT-PCR has been linked to documented false negatives and concerns over its reliability [3] especially when the test is done under compromised laboratory conditions. Chest computed tomography (CT) scans have been shown to have heightened sensitivity in diagnosing COVID-19 too, but their cost and technology implications are a trade-off when a large number of tests are required. [4].

On the contrary, scientific techniques are employed to identify the disease in its advanced phases by measuring antibody reactions [5]. Currently, commercially available fast antigen and molecular testing provides prompt results and offers several benefits, including reduced patient mortality rates and decreased hospital expenditures. Nevertheless, it is worth noting that these exams present challenges for individuals needing more specialized knowledge in the respective industry and failing to adhere to environmentally friendly principles [6]. In addition, the healthcare sector has extensively employed AI to analyze respiratory recorded files and identify diseases even before become aparent [7], [8].

The recent surge in leveraging cough sounds for diagnostic purposes has been notably documented [9]. For instance, Orlandic et al. highlighted the diagnostic potential of cough audio signals, emphasizing the role of machine learning in COVID-19 screening using the COUGHVID dataset [10]. However, the journey from raw cough sounds to actionable insights is not straightforward. Zhang et al. explored the challenges of continuous audio-based cough detection, especially concerning power consumption in wearables [11]. This underscores the need for efficient representation and processing of cough sounds. Enter the world of embeddings, where complex audio characteristics are distilled into compact vector forms leveraging the prowess of the Wav2Vec2.0 model [12]. But with high-dimensionality comes the challenge of efficient processing, making Techniques like Principal Component Analysis (PCA) have proven instrumental in mitigating these challenges, ensuring data integrity while streamlining processing.

The landscape of cough sound analysis is vast [13] delved into the segmentation of cough sounds, emphasizing the importance of segregating multiple coughs in a waveform into individual entities [13]. Such preprocessing steps are crucial for any subsequent analysis, especially when employing advanced techniques like graph-based semi-supervised learning. The potential of graph-based methodologies in this domain is vast, yet largely uncharted. Drawing inspiration from the world of social networks, where the connections between entities (or nodes) can reveal underlying patterns.

The "AudioVec-Diagnosis" framework is a new way to diagnose health problems by analyzing the full range of cough sounds that can be seen in **Fig. 1** which utilizes machine learning and deep learning algorithms to improve understanding of disease classification based on cough audio samples. Under the hood it uses the Wav2Vec2.0 model to convert cough sounds into vector embeddings that effectively capture complex audio features. PCA was employed to handle large dimensionality of the embeddings effectively.
The challenges we address in this paper goes into more detail about the problems that come up with high-dimensional audio embeddings and gives useful insights into how well different models can classify respiratory diseases using sound files as follow: **C1:** What is the best-performing machine learning model in a very few labeled(l) and large unlabeled (u) data points ($l << u$)? **C2:** Are the best performing models significantly different?

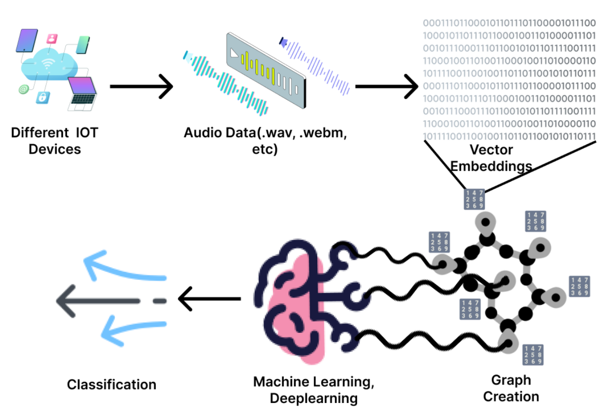

Fig. 1: Overview of Disease Classification Framework

II. Related Work

According to Bales et al.'s research, cough sounds contain a significant amount of diagnostic information. This claim provides a basis for the argument that machine learning can identify and diagnose medical conditions by analyzing these cough sounds [14] [18]. Mansour et al. demonstrated the integration of deep learning techniques with the Internet of Things (IoT) for the diagnostic analysis of tongue colour pictures in response to the increasing prevalence of IoT devices collecting health-related audio data [15]. Historically, the categorization of coughs has predominantly relied on utilizing fully supervised machine-learning methodologies [16] [17] [18]. Nevertheless, despite its considerable potential, semi-supervised learning (SSL) has received limited attention [19]. Using unlabeled data by SSL enhances the classification performance and addresses the difficulties arising from the limited availability of labeled datasets [20] [21]. The research by Anderson et al., which emphasizes the potential advantages of utilizing the vast amount of available unannotated data [22], has supported the use of unlabeled data. Semi-supervised models have demonstrated advantages over fully supervised counterparts in domains such as audio signal classification and cough recognition [21] [23]. Han et al. emphasized the potential benefits of incorporating SSL (Semi-Supervised Learning) into sound classification. They found that this integration may significantly decrease the human annotation required by 52.2%. Furthermore, their results demonstrated that SSL models could achieve performance levels equivalent to fully supervised models [24].

One of the notable SSL approaches is Pseudo-Label, which functions by initially training a model using labeled data. Pseudo-labels are a result of the subsequent categorization of unlabeled data, which is based on projected class probabilities [25]. SSL has played a crucial role in effectively dealing with the difficulties presented by inconsistent or ambiguous labels in classification efforts [26] [27]. In the field of 3D image segmentation, semi-supervised learning (SSL) models have shown that they can handle human labelling mistakes and noise well, often outperforming fully supervised models [25]. Zhu et al. have suggested an iterative semi-supervised learning (SSL) approach that leverages ambiguous emotion labels in speech emotion recognition. The results of their study indicate that commencing with a significant quantity of training data, even if the labels are relatively accurate, might improve classification outcomes compared to fully supervised models [28]. In the realm of SSL, notable advancements have been made, particularly in reconciling expert medical labels. These advancements involve utilizing a technique whereby a Deep Neural Network represents each expert. A final label is assigned to each sample by combining the findings from many different models [29]–[32]. Li et al. (2021) utilized the abovementioned strategy in their electronic medical record entity identification study. They applied the technique of Pseudo-Label to augment the dataset and determined the final labels using a majority vote [33]. Guan et al. demonstrated the superiority of a comparable approach in classifying diabetic retinopathy, as opposed to conventional approaches such as the Expectation-Maximization method, which is often utilized for assessing the accuracy of multiple raters [34] [35]. However, this sophisticated SSL method has yet to be explored in the context of audio signal categorization.

In the context of audio signal processing, Mari and Salarian [36] have proposed that the conversion of one-dimensional audio waveforms into two-dimensional representations necessitates the utilisation of novel methodologies capable of effectively managing the complex relationships inherent in the data. The endeavour to achieve proficient translation and comprehension of a wide range of audio signals has prompted the investigation of space filling curves (SFCs). Madani et al. employed a novel approach combining supervised and semi-supervised deep learning methodologies to diagnose heart conditions using echocardiography data. They highlighted the need to convert audio samples into condensed vector embeddings [37]. Dimensionality reduction techniques, such as Principal Component Analysis (PCA), play a crucial role in addressing the difficulties posed by high-dimensional embeddings. In their study, Liu et al. (2022) employed a generative adversarial network combined with pseudo-labeling to perform medical picture classification [38]. The authors emphasized the importance of principal component analysis (PCA) in their approach. Graph-based methodologies have demonstrated significant efficacy in the extraction of patterns. Huang and Chung (2020) proposed a disease diagnosis approach that use graph convolutional neural networks, which is in line with the suggestion made by AudioVec-detection to utilize Graph Neural Networks (GNNs) [39]. Purpura-Pontoniere et al. (2023) underscored the potential of graph-based methodologies in the field of audio analysis [40]. In summary and from the studied related work, while traditional fully supervised strategies have proved be crucial in cough classification, merging semi-supervised methodologies and advanced processing techniques presents a promising trajectory for future and as viable alternatives in audio-based chronic disease detection.

III. Methods

A. Dataset Description

The Coswara and COUGHVID datasets provide a comprehensive collection of respiratory audio recordings that are important to study and detect SARS-CoV-2 conditions. The Coswara dataset comprises a total of 4,701 audio recordings collected over the period from April 2020 to February 2022. Among these recordings, 1,182 came from people who tested positive for COVID-19, while the remaining 3,519 were acquired from healthy individuals. The dataset contains nine separate kinds of respiratory sounds, accompanied by comprehensive metadata, and predominantly obtained from India, but we are considering only cough. Simultaneously, the COUGHVID crowdsourcing dataset provides 9,304 audio samples, comprising 1,092 recordings from individuals who tested positive for COVID-19 and 8,212 recordings from individuals in good health. These datasets were combined to provide a strong basis for researchers seeking to utilize audio analytics for the early diagnosis of COVID-19 and maybe other respiratory diseases.

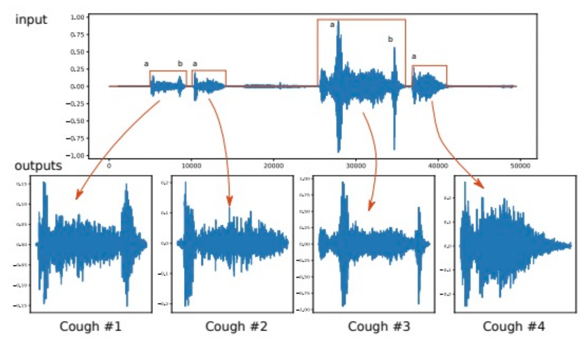

Fig. 2: Cough Audio Segment

B. Data Preprocessing

In the data preprocessing stage, the audio recordings are initially converted from the .webm format to the more widely recognized .wav format. The recordings were subsequently subjected to a systematic categorization process, wherein designations such as ”symptomatic,” ”COVID-19,” and ”healthy”. After an initial data investigation, thorough processing was applied to these audio recordings. The cough sample data underwent downsampling, standardization, and filtering processes. Then several separate cough segments were separated from a single audio file **Fig. 2**. The aggregated information played a crucial role in differentiating between several categories, namely those who were in excellent health, those who were displaying symptoms, and those who had received a confirmed diagnosis of COVID-19. As a result of the scarcity of COVID-19 files, the audio recordings were divided into discrete intervals of 5 seconds each. Upon completing the processing stage, the data from both datasets, namely the COVID-19 and healthy datasets, were gathered and systematically arranged into their respective folders.

The third phase of the workflow was the extraction of feature vectors from the segmented audio files contained within the folders labeled ”COVID” and ”healthy.” We Employed the Facebook Wav2Vec 2.0 model., which has received recognition for its exceptional performance in several audio-oriented applications. The model successfully transformed unprocessed audio data into vector embeddings using a function explicitly created for this purpose. The embeddings were subsequently averaged, generating a singular representative embedding vector for every audio sample as seen in **Fig. 3**. Ultimately, the embeddings obtained from the COVID-19 and healthy categories were added into a unified data frame. The data frame underwent a transformation procedure in which the list of embeddings was enlarged into separate columns, facilitating a comprehensive examination.

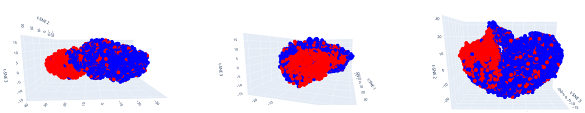

Fig. 3: 3D View of the Data Points

After performing feature extraction, the embeddings obtained from both the COVID and healthy categories were combined into a unified DataFrame. The DataFrame underwent a transformation procedure in which the list of embeddings was enlarged into separate columns, preparing it for detailed examination. PCA was utilized as a dimensionality reduction technique due to the inherent high-dimensionality of the embeddings. The application of this methodology resulted in the reduction of the embeddings to a set of five primary components. This approach guarantees the preservation of essential information while simultaneously simplifying the data structure. The dataset, which has been optimized by reducing its dimensions, was divided into separate subsets for training and testing purposes. The strategic split ensured that the models were trained on a certain portion of data and then evaluated on a separate subset of data that had not been previously viewed.

C. Experiments

The models chosen during the analysis are Logistic Regression, Random Forest, XGBoost, K-Nearest Neighbours (KNN), Label Spreading, and Label Propagation. This gave us important information about how they behave and how well they work. Also, two well-known deep learning models, the Graph Convolutional Network (GCN) and the Graph Attention Network (GAT), were tested to see how well they work with non-linear data points. In order to investigate **C1**, the dataset was subjected to a systematic partitioning process and divided into predetermined proportions (1%, 3%, and 5%) for training and testing. This segmentation aimed to assess the performance of models when exposed to different levels of training data availability. Following the partitioning process, the models underwent training using the designated training datasets and then evaluation using the testing datasets. During the training process, the models tried to comprehend the complex patterns and interconnections in the data, strengthening their skills to make accurate predictions. The performance of the models was evaluated using a comprehensive set of metrics, including accuracy, precision, recall, F1-score, receiver operating characteristic (ROC) area under the curve (AUC), and precision-recall (PR) AUC. The combination of these indicators provided a comprehensive view of the models' prediction performance and ability to handle skewed data distributions effectively.

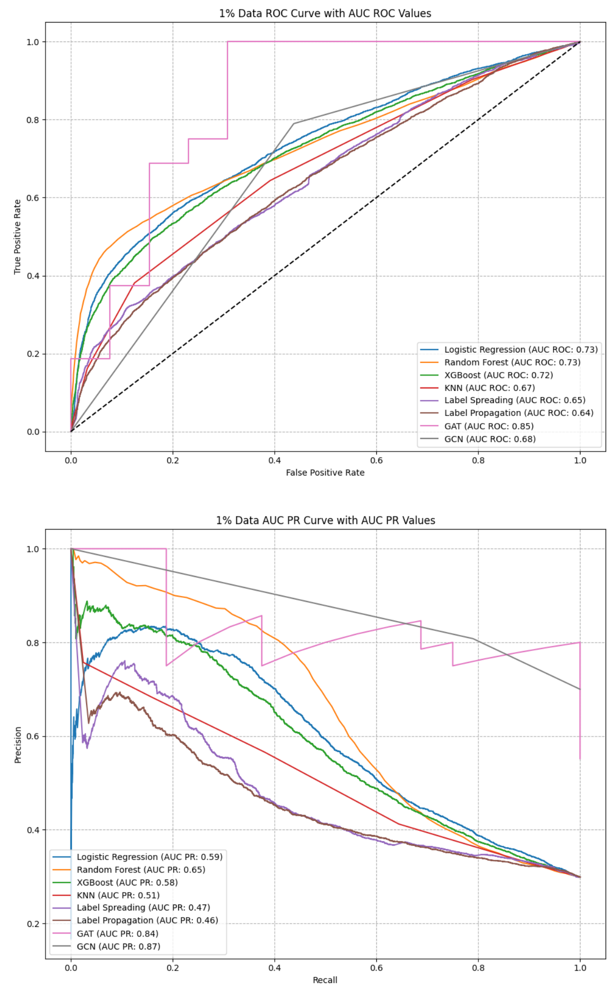

Fig. 4: 1% AUC PR & ROC Curve

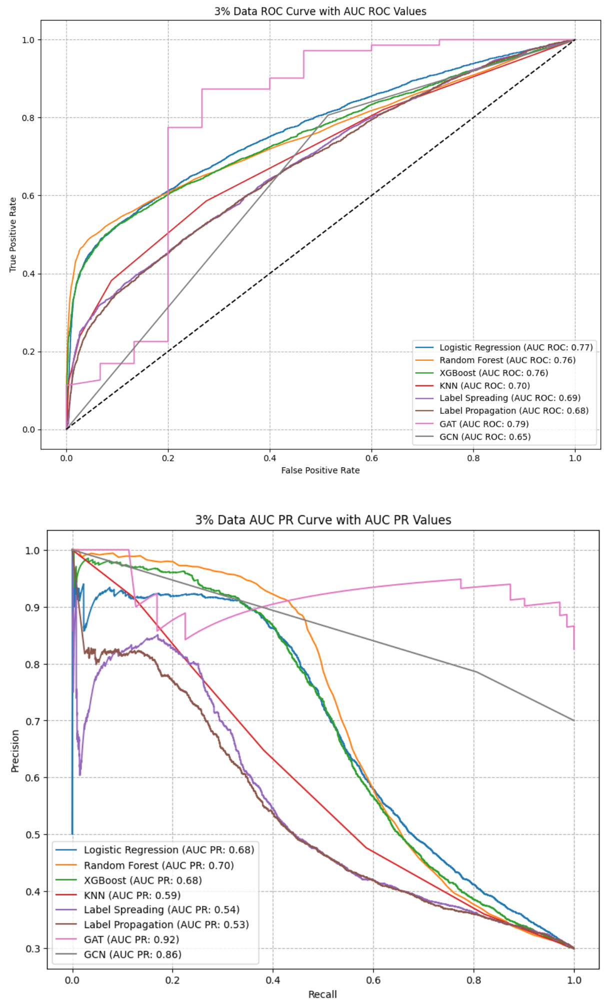

Fig. 5: 3% AUC PR & ROC Curve

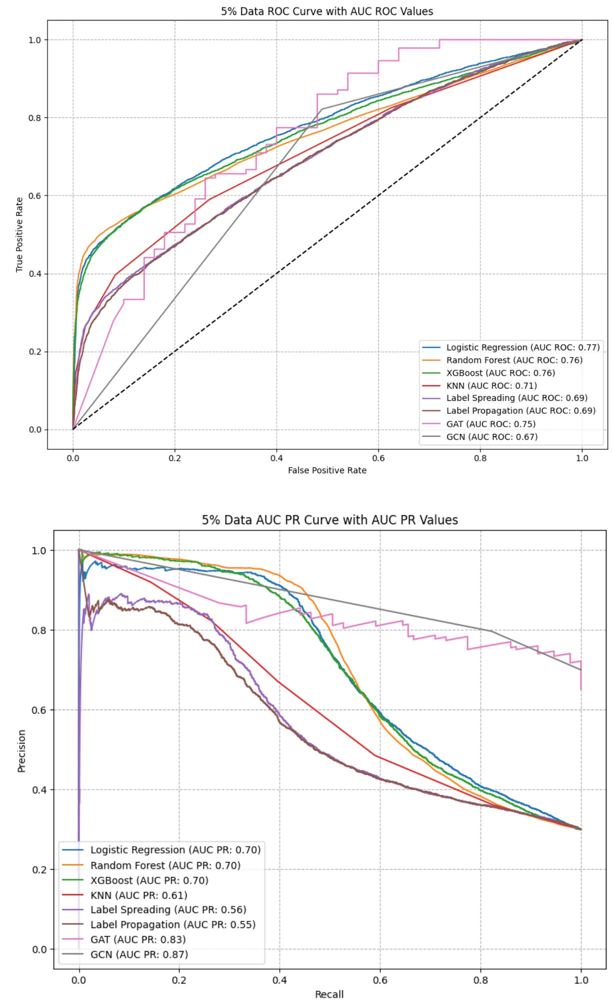

Fig. 6: 5% AUC PR & ROC Curve

In order to enhance comprehension of the models' performance, **Fig. 4, 5, 6** ROC and PR curves were generated by altering decision criteria. The visual depictions clearly understood the relationship between several parameters, including true positive rates, false positive rates, precision, and recall. In addition, to facilitate future comparative analysis, data serialization techniques were utilized to store each model's calculated ROC and PR curve values across varying dataset sizes. **C2** was addressed comparative analysis assessed the performance indicators of the two predominant models. The primary focus was given to ROC and PR metrics since they hold significant relevance in evaluating the efficacy of models when dealing with imbalanced datasets. The AUC values for the 1%, 3%, and 5% data samples were subjected to a paired t-test. The resulting p-values for the comparisons between the top models were determined to be $p = 0.926$ for AUC PR (GAT vs. GCN) and $p = 0.438$ for AUC ROC (GAT vs. Logistic Regression). Based on the standard alpha threshold of 0.05, neither calculated p-values can be considered statistically significant. According to the data points analyzed, there is insufficient evidence to reject the null hypothesis that the models exhibit comparable performance in terms of AUC PR and AUC ROC. Nevertheless, it is imperative to acknowledge the constraints associated with this assessment, namely the use of merely three data points per model and the non-independence of these data points, as they stem from varying proportions of the same dataset. Considering the constraints mentioned above, it is advisable to exercise caution when interpreting the resulting p-values. However, they offer a first understanding of the relative performances of the models.

IV. Discussions

The quantity and quality of data significantly impact the effectiveness of models in machine learning. This paper presents empirical findings regarding the impact of data volume on the performance of both traditional and graph-based models, specifically 1%, 3%, and 5% of the overall dataset. Previous research has shown an increase between increased data volume and enhanced performance across various models.

The results reveals that as the data volume increases, most models have a improvement that can be seen in performance metrics. Traditional machine learning models, such as Random Forest and Logistic Regression, show a consistent upward move in their AUC PR and ROC values with the increase in data size. On the other hand, graph-based models like GAT and GCN show a more complex success trajectory, which shows how sensitive they are to the details of the data. For traditional machine learning models, Table I metrics like precision, recall, and accuracy (as presented in the tables) indicate that Random Forest often surpasses other models' precision and accuracy. However, Table II the graph-based GAT model excels in recall, suggesting its proficiency in correctly identifying a higher proportion of positive cases. When graph-based models are compared with standard ML models, exciting things can be learned. Models like Random Forest and Logistic Regression do well across many measures, but graph-based models, especially GCN, do very well in the AUC PR metric.

Model	Prec.	Rec.	Acc.	ROC	PR
RF	**0.8126**	0.4943	**0.8147**	**0.7747**	0.7037
LR	0.7634	0.4911	0.8023	0.7747	0.6984
XGB	0.7359	0.5037	0.7975	0.7640	0.6974
KNN	0.6697	0.3966	0.7610	0.7095	0.6099
LS	0.5252	0.4495	0.7138	0.6942	0.5640
LP	0.5153	0.4552	0.7090	0.6912	0.5549

TABLE I: Metrics for Traditional ML Models

Model	Prec.	Rec.	Acc.	ROC	PR
GAT	0.7395	**0.9462**	0.7483	0.7503	0.8349
GCN	0.7958	0.8217	0.7280	0.6662	**0.8710**

TABLE II: Metrics for Graph-based Models

V. Conclusion and Future Work

The paper extensively explores the significance of Supervised learning and semi-supervised learning. Significant insights have been obtained through rigorous experimentation involving machine learning techniques and graph-based models. Traditional machine learning models, such as Random Forest and Logistic Regression, have repeatedly demonstrated robust performance across diverse datasets of varied sizes. On the other hand, graph-based models, particularly the Graph Convolutional Network (GCN), have proven how well they uncover latent relationships inside complex audio embeddings. With changed training data size (1%, 3%, 5%) for the algorithm, the experiment design has shown a clear relationship between success metrics and data volume. The outcomes of the study emphasize the significance of both the quantity and quality of data in assessing the effectiveness of models. Moreover, the findings underscore the capacity of graph-based models in effectively managing complex data structures, particularly in cases when the dataset contains a wealth of interconnections and associations. Although the Graph Convolutional Network (GCN) and Graph Attention Network (GAT) have demonstrated promising results, exploring alternative advanced graph-based models to achieve even higher performance levels is worth considering. Further examination of the mis-classifications made by each model could offer valuable insights into their limitations and potential avenues for enhancement. Considering the demonstrated potential of semi-supervised learning, it might be advantageous to investigate alternative strategies within the realm of SSL for further precision.

Acknowledgment

This work is partially supported by the EU-Horizon Europe FAME 101092639 and The INSIGHT SFI research Centre for Data Analytics supported by Science Foundation Ireland Grant Number SFI/12/RC/2289.

References

[1] A. Atalan, "Is the lockdown important to prevent the covid-19 pandemic? effects on psychology, environment and economy-perspective," *Annals of Medicine and Surgery*, vol. 56, pp. 38–42, 8 2020.

[2] K. Inoue, S. Hashioka, and N. Kawano, "Risk of an increase in suicide rates associated with economic downturn due to covid-19 pandemic," *Asia-Pacific journal of public health*, vol. 32, p. 367, 9 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32639165/

[3] Y. Li, L. Yao, J. Li, L. Chen, Y. Song, Z. Cai, and C. Yang, "Stability issues of rt-pcr testing of sars-cov-2 for hospitalized patients clinically diagnosed with covid-19," *Journal of medical virology*, vol. 92, pp. 903–908, 7 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32219885/

[4] Y. W. Tang, J. E. Schmitz, D. H. Persing, and C. W. Stratton, "Laboratory diagnosis of covid-19: Current issues and challenges," *Journal of clinical microbiology*, vol. 58, 6 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32245835/

[5] J. Dinnes, J. J. Deeks, A. Adriano, S. Berhane, C. Davenport, S. Dittrich, D. Emperador, Y. Takwoingi, J. Cunningham, S. Beese, J. Dretzke, L. F. di Ruffano, I. M. Harris, M. J. Price, S. Taylor-Phillips, L. Hooft, M. M. Leeflang, R. Spijker, and A. V. den Bruel, *The Cochrane database of systematic reviews*.

[6] K. C. Santosh, "Speech processing in healthcare: Can we integrate?" *Intelligent Speech Signal Processing*, pp. 1–4, 1 2019.

[7] H. Mukherjee, P. Sreerama, A. Dhar, S. M. Obaidullah, K. Roy, M. Mahmud, and K. C. Santosh, "Automatic lung health screening using respiratory sounds," *Journal of Medical Systems*, vol. 45, pp. 1–9, 2 2021. [Online]. Available: https://link.springer.com/article/10.1007/s10916-020-01681-9

[8] H. Mukherjee, H. Salam, and K. Santosh, "Lung health analysis: Adventitious respiratory sound classification using filterbank energies," *https://doi.org/10.1142/S0218001421570081*, vol. 35, 11 2021.

[9] S. Yao, R. Wang, S. Sun, D. Bu, and J. Liu, "Joint embedding learning of educational knowledge graphs," pp. 209–224, 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-41099-5_12

[10] L. Orlandic, T. Teijeiro, and D. Atienza, "The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Scientific Data 2021 8:1*, vol. 8, pp. 1–10, 6 2021. [Online]. Available: https://www.nature.com/articles/s41597-021-00937-4

[11] S. Zhang, E. Nemati, M. Dinh, N. Folkman, T. Ahmed, M. Rahman, J. Kuang, N. Alshurafa, and A. Gao, "Coughtrigger: Earbuds imu based cough detection activator using an energy-efficient sensitivity-prioritized time series classifier," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2022-May, pp. 1–5, 2022.

[12] T. S. Nguyen, J. Lee, V. T. Hoang, and O.-J. Lee, "Connector 0.5: A unified framework for graph representation learning," 4 2023. [Online]. Available: https://arxiv.org/abs/2304.13195v1

[13] B. T. Atmaja, A. S. AIST, and K. K. K. H. S. G. S. Kenkyujo, "Evaluation of automatic single cough segmentations," 1 2023. [Online]. Available: https://europepmc.org/article/ppr/ppr594628

[14] C. Bales, M. Nabeel, C. N. John, U. Masood, H. N. Qureshi, H. Farooq, I. Posokhova, and A. Imran, "Can machine learning be used to recognize and diagnose coughs?" *2020 8th E-Health and Bioengineering Conference, EHB 2020*, 4 2020. [Online]. Available: http://arxiv.org/abs/2004.01495http://dx.doi.org/10.1109/EHB50910.2020.9280115

[15] R. F. Mansour, M. M. Althobaiti, and A. A. Ashour, "Special section on deep learning for internet of things internet of things and synergic deep learning based biomedical tongue color image analysis for disease diagnosis and classification."

[16] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "A cough-based algorithm for automatic diagnosis of pertussis," *PloS one*, vol. 11, 9 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27583523/

[17] T. Xia, D. Spathis, C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, E. Bondareva, T. Dang, A. Floto, P. Cicuta, and C. Mascolo, "Covid-19 sounds: A large-scale audio dataset for digital respiratory screening." [Online]. Available: https://covid19.who.int/

[18] J. Laguarta, F. Hueto, and B. Subirana, "Covid-19 artificial intelligence diagnosis using only cough recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 275–281, 2020.

[19] Z. Zhu and Y. Sato, "Speech emotion recognition using semi-supervised learning with efficient labeling strategies," *2021 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2021 - Proceedings*, pp. 358–365, 2021.

[20] DengJun, XuXinzhou, ZhangZixing, FruhholzSascha, Schuller-Bjorn, J. Deng, X. Xu, Z. Zhang, FruhholzSascha, and SchullerBjorn, "Semisupervised autoencoders for speech emotion recognition," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, pp. 31–43, 1 2018. [Online]. Available: https://dl.acm.org/doi/10.1109/TASLP.2017.2759338

[21] M. Y. Guan, V. Gulshan, A. M. Dai, and G. E. Hinton, "Who said what: Modeling individual labelers improves classification," *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pp. 3109–3118, 2018.

[22] T. H. Huynh, V. A. Tran, and H. D. Tran, "Semi-supervised tree support vector machine for online cough recognition," *Interspeech 2011*, 2011. [Online]. Available: https://www.academia.edu/1057590/Semi_Supervised_Tree_Support_Vector_Machine_for_Online_Cough_Recognition

[23] M. Raghu, K. Blumer, R. Sayres, Z. Obermeyer, R. Kleinberg, S. Mullainathan, and J. Kleinberg, "Direct uncertainty prediction for medical second opinions," pp. 5281–5290, 5 2019. [Online]. Available: https://proceedings.mlr.press/v97/raghu19a.html

[24] R. Tanno, A. Saeedi, S. Sankaranarayanan, D. C. Alexander, and N. Silberman, "Learning from noisy labels by regularized estimation of annotator confusion."

[25] D.-H. Lee, "Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks," 2013.

[26] H. Xue and F. D. Salim, "Exploring self-supervised representation ensembles for covid-19 cough classification," *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 1944–1952, 8 2021.

[27] W. Yang, J. Zhao, Y. Qiang, X. Yang, Y. Dong, Q. Du, G. Shi, and M. B. Zia, "Dscgans: Integrate domain knowledge in training dual-path semi-supervised conditional generative adversarial networks and s3vm for ultrasonography thyroid nodules classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11767 LNCS, pp. 558–566, 2019.

[28] Z. Zhang and B. Schuller, "Semi-supervised learning helps in sound event classification," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 333–336, 2012.

[29] W. Han, E. Coutinho, H. Ruan, H. Li, B. Schuller, X. Yu, and X. Zhu, "Semi-supervised active learning for sound classification in hybrid learning environments," *PloS one*, vol. 11, 9 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27627768/

[30] Z. H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, pp. 44–53, 1 2018. [Online]. Available: https://dx.doi.org/10.1093/nsr/nwx106

[31] J. Lv, X. Chen, J. Huangy, and H. Bao, "Semi-supervised mesh segmentation and labeling," *Computer Graphics Forum*, vol. 31, pp. 2241–2248, 9 2012. [Online]. Available: https://dl.acm.org/doi/10.1111/j.1467-8659.2012.03217.x

[32] R. Shi, L. Wang, and J. Jiang, "An unsupervised region of interest extraction model for tau pet images and its application in the diagnosis of alzheimer's disease," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2022, pp. 2157–2160, 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/36083928/

[33] Z. Li, Z. Gan, B. Zhang, Y. Chen, J. Wan, and J. Zhao, "Noisy label learning for chinese medical named entity recognition based on uncertainty strategy."

[34] A. P. Dawid and A. M. Skene, "Maximum likelihood estimation of observer error-rates using the em algorithm," *Applied Statistics*, vol. 28, p. 20, 1979.

[35] S. Kumar, M. K. Chaube, S. H. Alsamhi, S. K. Gupta, M. Guizani, R. Gravina, and G. Fortino, "A novel multimodal fusion framework for early diagnosis and accurate classification of covid-19 patients using x-ray images and speech signal processing techniques," *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107109, 11 2022.

[36] A. Mari and A. Salarian, "A novel audio representation using space filling curves," *ArXiv*, 2022.

[37] A. Madani, J. R. Ong, A. Tibrewal, and M. R. Mofrad, "Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease," *npj Digital Medicine 2018 1:1*, vol. 1, pp. 1–11, 10 2018. [Online]. Available: https://www.nature.com/articles/s41746-018-0065-x

[38] K. Liu, X. Ning, and S. Liu, "Medical image classification based on semi-supervised generative adversarial network and pseudo-labelling," *Sensors 2022, Vol. 22, Page 9967*, vol. 22, p. 9967, 12 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/24/9967/htmhttps://www.mdpi.com/1424-8220/22/24/9967

[39] Y. Huang and A. C. Chung, "Semi-supervised multimodality learning with graph convolutional neural networks for disease diagnosis," *2020 IEEE International Conference on Image Processing (ICIP)*, vol. 2020-October, pp. 2451–2455, 10 2020.

[40] A. Purpura-Pontoniere, D. Terzopoulos, A. Wang, and A.-A.-Z. Imran, "Semi-supervised relational contrastive learning."