Improving Automated Medical Transcriptions with Few-Shot Learning

Dr. Ashima Yadav
School of Computer Science Engineering and Technology
Bennett University
Greater Noida, India
ashimayadavdtu@gmail.com

Nimratpreet
School of Computer Science Engineering and Technology
Bennett University
Greater Noida, India
knimrat345@gmail.com

Sahil Khandekar
School of Computer Science Engineering and Technology
Bennett University
Greater Noida, India
sahilasif500@gmail.con

Anannya Pandey
School of Computer Science Engineering and Technology
Bennett University
Greater Noida, India
anannyapandey08@gmail.com

***Abstract*—Automated medical transcription systems have made eminent strides in recent years, yet challenges remain in faultlessly transcribing medical terminology, varied accents, and specialized language with inadequate labelled data. Traditional transcription systems depend predominantly on enormous, domain-specific datasets, that are often high priced and too time-consuming to acquire. Few-shot learning (FSL), is a paradigm that allows the models to generalize from less number of examples, offering a very optimistic solution to this problem. This paper explores the implementation of few-shot learning processes to intensify the performance of automated medical transcription systems, particularly in scenarios with scant data availability for training. We suggest a framework that combines pre-trained language models, meta-learning, and transfer learning to refine transcription accuracy with nominal labelled examples. Investigational outcomes demonstrate that our few-shot approach outperforms traditional models, achieving a reduction in word error rate and better grasping many intricate medical terms. Our outcomes suggest that few-shot learning has the capability to revolutionize transcriptions of medical domain, offering a prestigious and scalable solution for applications in healthcare.**

Keywords*—*Automated Medical Transcription, Few-Shot Learning (FSL), Medical Terminologies, Pre-trained Language Models, Meta-Learning, Transfer Learning, Word Error Rate.

I. Introduction

In the healthcare sector, accuracy and efficiency of transcription of medical information is considerably essential for maintaining patient records, facilitation of communication amongst the healthcare executives, and corroborating proper clinical documentation. Few-shot learning is also called low-shot learning. It is a set of machine learning techniques that try to learn executing tasks using small amounts of labelled training instances[1]. Medical transcription alludes to the procedure of conversion of spoken or audio-recorded medical information into written text. This transcription acts as the chief means of documenting professional- patient interactions, diagnoses, treatment charts, and clinical observations. With the increasing clamor for healthcare services, there is a constantly increasing need for automated transcription systems that are accurate and hold the potential of efficiently converting spoken medical language into text, without encumbering healthcare professionals with tasks related to manual transcription.

However, this task of automating medical transcription doesn't come without challenges. Implementation of the in-context learning capability in a medical background is challenging because of the intrinsic complications and multimodality of data in medical domain and the diverse tasks to be solved [2]. The medical field is characterized by specialized expressions, complex jargon, and an extensive array of abbreviations and medical codes. These terminologies can vary not only across dissimilar medical specialties but also within divergent regions and among individual healthcare providers. The language used in medical setups is often informal and filled with unprompted speech patterns, disfluencies, and sometimes even dialectal differences. Furthermore, transcription systems must contend with a vast range of environmental noises, such as chatter in the background or sounds of medical equipment , which can intercede in the clarity of the spoken language.

Conventional automated transcription systems depend on massive, annotated datasets to learn the patterns of speech and language needed for precise transcription. These datasets typically contain thousands or even millions of hours of labelled audio paired or converted into corresponding transcriptions. However, the collection and annotation of such substantial data for the medical domain confers significant challenges. Attaining high-quality, domain-specific labelled datasets is both time-taking and expensive. In the medical domain, especially in extremely specialized regions or for very rare medical conditions, the data needed for training transcription models is often bounded. This sparseness of data creates a significant roadblock for developing high performance transcription that can handle the vast and diverse vocabulary and varied speech patterns typical in healthcare settings.

Few-shot learning is a subfield of machine learning that presents a favorable solution to this problem. Few-shot applications enters the stage as a valuable connection between fully supervised and zero-shot methods [3]. Unlike traditional approaches conferring machine learning models, which require enormous labelled data to perform efficiently, few-shot learning techniques are delineated to enable the models to gain knowledge from a minute quantity of labelled instances. Its capability lies in its proficiency to generalize from just a handful of training data, making it a structured and exemplary approach for sectors where labelled data is less or challenging to attain. In the domain of medical imaging, due to the infrequent occurrence of some diseases, there is often a restriction on the available data, as an outcome, of which the success of few shot learning algorithms can prove to be an eminent evolution[4].

Few-shot learning is assembled on the notions of meta-learning, which lets the models learn how to adjust expeditiously to new tasks or domains with less data. Few-shot learning is a machine learning technique for learning a task from bounded data, images and guidance[5]. In the factors of medical transcription, a few-shot learning model could learn to recognize and accurately transcribe medical terminology, even with only a few examples. Furthermore, transfer learning, another technique is commonly used in few-shot learning, which enables the models to influence knowledge inculcated from one field to improve execution in the other. In this case, language models that are pre trained can be finely tuned on a small, domain-centric medical dataset to help the system adjust to the nuances of medical language. By combining these few-shot learning methods, we can develop a system that not only lowers down the dependence on massive, labelled datasets but also enhances the performance and efficiency of automated medical transcription. The usage of pre-trained models that already has general knowledge of language allows the system to get used to the specific needs of medical transcription. Meta-learning validates the system to finely adapt to new medical terms or dialects with limited examples, while transfer learning facilitates the system to apply knowledge from other sectors, improving its overall accuracy even with small quantities of medical data.

The application of few-shot learning in automated medical transcription could have prominent implications for the healthcare industry. FSL is very crucial for clinical NLP as annotating a large training data set is very expensive and usually requires involving domain experts [6] .It would allow the growth of transcription systems that are more adaptable and can be deployed across a great range of medical specialties, even those with restricted training data. Additionally, it could follow to transcription systems that are more useful and cheaper, reducing the need for extensive data annotation efforts. By allowing the creation of extremely accurate transcription systems with fewer labelled examples, few-shot learning could eventually reduce the administrative burden on healthcare professionals, allowing them to focus prominently on patient care rather than the time-consuming work of documentation. The algorithms and implementation of Few-shot learning have evolved and have proven to be a significant tool in scenarios with small data [7]. Since, It is not practical to train a fully supervised new model for every new class where segmentation occurs, training from the very beginning or fine tuning are time-consuming and they require expertise [8] which is why the need for few shot learning occurred.

This paper explores the features and abilities of few-shot learning techniques to enhance automated medical transcription systems, particularly in situations where data is limited. We intend to propose a novel framework that connects pre-trained language models, meta-learning, and transfer learning to enable transcription systems for accurate performance with nominal labelled data. Our approach aims to address the challenges of medical transcription system, such as the need for specialized medical terminology, regional accents, and noisy environments, by allowing the system to generalize from just a few examples. Through extensive experimentation and evaluation, we aim to exhibit that few-shot learning can provide noteworthy improvements in transcription accuracy, especially in cases where traditional models struggle because of data scarcity or domain-specific challenges.

II. Related Work

The task of medical transcription has long been a crucial factor of healthcare documentation, ensuring that precise patient records are maintained for clinical decision-making and legal purposes. Training high-performing models require enormous amounts of pixel-level ground truth masks, which can be prohibitive to achieve in the medical domain [9]. The evolution of transcription technologies has moved from human-made manual efforts, to small scale and large scale processes being assisted by machines. Automated medical transcription systems hold the capacity to revolutionize healthcare by significantly lessening the time and cost associated with manual transcription, as well as performing improvements in the accuracy of transcribed data. However, the complexity and specificity of medical language, along with challenges such as accent variation, diverging speech structures, and noise in healthcare environments, continue to make fully automated transcription a difficult task. In retaliation to these challenges, researchers have surveyed various machine learning (ML) techniques, some of these includes transfer learning, supervised learning, few shot learning, and meta learning, to improve medical transcription systems. Few shot learning is developed to alleviate this burden, which attains competitive performances with only some labelled data [10].

This section concerns with reviewing the key approaches and techniques applied to automated medical transcription, the challenges faced in the field, and how recently new developments in few-shot learning and other methods related to it have been used to address these challenges.

A. *Early Approaches to Medical Transcription*

Earlier, medical transcription used to rely profoundly on manual processes, where medical professionals or dedicated transcriptionists used to listen to audio recordings of medical dialogues and typed the content into text. This process, though meticulous, was time-consuming, prone to mistakes, and not efficient, especially while dealing with gigantic volumes of information. With recognition of Automatic Speech Recognition (ASR) in the years 1980s and 1990s, efforts for the automation of medical transcription gained momentum. The goal of ASR systems was to convert spoken language into text by using phonetic, linguistic, and statistical models to match speech to corresponding written words. Historically, these systems were restricted in their aptness to handle medical speech due to the challenges posed by the specialized nature of medical language, accompanied by the utilization of abbreviations, acronyms, and the complexity of accents and dialects. These systems were trained principally on generalized speech data, often leading to enhanced error rates when applied to medical transcription tasks. Medical ASR systems also struggled with background noise, disfluencies in speech, and the speedy pace of medical interactions, resulting in less accuracy when compared to general speech recognition systems. The medical terms were fine-grained, making them arduous to recognize[11].

B. *Advancements in Machine Learning for Medical Transcription*

The limitations of early ASR systems motivated researchers to explore more sophisticated machine learning techniques. Machine learning, supervised learning to be precise, surfaced as a major approach in addressing the challenges of medical transcription. These methods use labelled data for model training, with the aim of learning structures and relations connecting the speech capture and text corresponding output. Large-scale annotated datasets which includes clinical narratives and medical transcripts, have been instrumental in training and evaluation of medical transcription systems.

Recently, deep learning techniques , such as RNNs or recurrent neural networks and long short-term memory or LSTM networks, have been utilized to perform medical transcription tasks. These models, which excel at handling sequential data, have suggestively improved the accuracy of transcription systems. Also, the introduction of CNNs or convolutional neural networks used for feature extraction and transformer based architectures, such as BERT and GPT, has further made advancements in transcription systems by allowing them to capture the contextual and semantic relationships between medical terms in a much better way.

While deep learning techniques have led to substantial improvements in transcription accuracy, there are still several challenges to face in the medical domain. A prerequisite for deep learning is large amounts of labelled data, collection of which is very difficult, costly, time-consuming and lengthy for applications of crucial nature like the medical field [12]. Attaining sufficient labelled data for rare medical conditions, uncommon procedures, and specialized medical terminology remains an important hurdle in developing robust transcription systems. Plus, the great region gist existing in analysis in medical-field from inter-stain, inter-population, or inter-scanner variability needs regular adaptation, most probably done in a data-efficient way by using minimal numbers of labelled instances, also known as few shot adaptations[13].

C. *Transfer Learning and Fine-Tuning for Medical Transcription*

Transfer learning is a method that involves leveraging pre-trained models on complicated, general datasets and adapting them to a specific domain or work with minor amounts of domain-specific data. This approach has been vastly applied in medical transcription to address the sparsity of labelled data. By performing fine-tuning on pre-trained models, researchers have been able to improvise transcription accuracy for medical terminology without the requirement for massive labelled datasets for every possible medical condition.

A noteworthy example of transfer learning in medical transcription is the use of BERT i.e. Bidirectional Encoder Representations from Transformers, that is a transformer based model that has been pre-trained on humongous amounts of common text data. Fine-tuning it on medical datasets has allowed researchers to significantly improves the understanding of model of medical domain, including recognizing complicated terms, medical abbreviations, and domain-specific jargon. Other models, such as BioBERT and ClinicalBERT, have been fine-tuned on medical corpora to increase their performance in clinical and medical text processing tasks.

While transfer learning has made substantial strides in improving medical transcription, it still faces a lot of challenges, particularly while dealing with highly specialized or low-resource medical domains. For example, transfer learning models trained on general medical data may still face issues to accurately transcribe rare or emerging medical conditions that are not well-represented in the training data. Moreover, fine-tuning large models requires significant computational resources, making it less accessible to healthcare providers with limited technical configurations.

D. *Meta Learning for Medical Transcription*

It is often also known as learning to learn and is a progressive machine learning paradigm that enables adaptation in models quickly to new tasks with lesser data. The objective of Meta learning algorithms is to learn the underlying structure of a task, allowing the model to transfer its knowledge to new tasks with bounded examples. These abilities make meta-learning a natural fit for improving medical transcription, especially in cases where data is scarce.

In medical transcription, meta-learning can allow models to quickly adapt to new medical domains or unique terminologies with only a few examples. Meta learning approaches, such as model-agnostic meta-learning also known as MAML, prototypical networks, and optimization-based methods, have been applied in wide ranges of natural language processing (NLP), predominantly text classification, named entity recognition or NER, and machine translation. These methods have shown promising results in improvising medical transcription systems by allowing them to generalize better to unconventional medical terms and conditions.

For example, a meta-learning approach could allow a transcription model to rapidly adjust to transcribing rare diseases or medical procedures with minimum examples. Meta-learning methods can also help transcription systems maneuver diverse accents, speech patterns, and dialects, which is a censorious aspect of medical transcription, as healthcare professionals come from varying linguistic backgrounds. The potential to quickly adapt to updated transcription tasks is particularly advantageous in healthcare environments where the medical landscape is constantly developing.

E. Few-Shot Learning in Medical Transcription

Few-shot learning (FSL), a subdivision of meta-learning, has received undivided attention in latest years because of its potential to address the matter of restricted mark data. Few-shot study allows models to carry out well on assignment with only a few categorized examples, creating it a perfect solution for healthcare transcriber, where large explain datasets may not be accessible for able to be done medical term, situation, or tone. Also, The few-shot learning [14,15,16] is suggested to solve the obstacle of overfitting and aims to acknowledge novel categories from very few labelled examples.

FSL approaches use techniques like metric learning, data augmentation, and meta training to enable models to generalize from a nominal quantity of instances. In the context of medical transcription, few-shot learning can be used for training models to recognize new medical terms, adapt to different speech patterns, and transcribe medical records accurately, even in low-resource settings. For example , few-shot learning has been used to find rare diseases or emerging medical terminology with only a few examples from clinical texts.

The applications of few shot learning in medical transcription is particularly favorable for handling domain-specific challenges, such as rare diseases, dialectal variations in speech, or variations in medical abbreviations. Few-shot learning approaches allow models to generalize across diversified medical contexts, improving transcription accuracy in specialized areas that lack enough training data. Additionally, FSL brings down the need for extensive data annotation, lessening the cost and time involved in developing magnificent medical transcription systems.

F. Recent Advances and Hybrid Approaches

Recent researches in medical transcription has focused mostly on blending various techniques, including transfer learning, deep learning, few-shot learning, and meta learning, to convey the challenges posed by medical language. Hybrid models that integrate multiple learning approaches have shown propitious results in improving transcription accuracy, especially in the areas of rare or specialized medical domains.

For instance, researchers have explored mixing pre-trained transformer frameworks with few shot learning techniques to build transcription systems that can work well even when only a minimum quantity of labelled data is available. By assimilating both transfer learning (to leverage general language knowledge) and few shot learning (for adaptation to specific medical domains), these hybrid models have shown refined performance in handling complex medical terminology and accents.

Additionally, recent studies have explored the use of reinforcement learning (RL) for fine-tuning transcription models, where the system iteratively increases its transcription performance by receiving feedback on its predictions. The blend of reinforcement learning with few-shot learning techniques could lead to systems that continuously adapt and improve as they face new transcription challenges in real-world clinical environments. Current advances have also shown the uses of FSL for text parsing [17], translation in machines [18], and classification [19], [20], amid others. Many utilization domains have also been researched for FSL in Natural Language Processing , such as law-related [21] and biomedical applications which are prominently the spotlight of this review.

Summing up, we can say that the methodology adopted to improve automated medical transcription using few-shot learning, with a focus on data preprocessing, model architecture, and evaluation. Given the difficulties associated with medical transcriptions, such as divergent terminology, varied accents, and limited labelled data, the goal of proposed approach is to leverage powerful deep learning techniques to increase transcription accuracy. We incorporate transformer based models and few shot learning, which enables the model to perform well with only a small quantity of labelled instances.

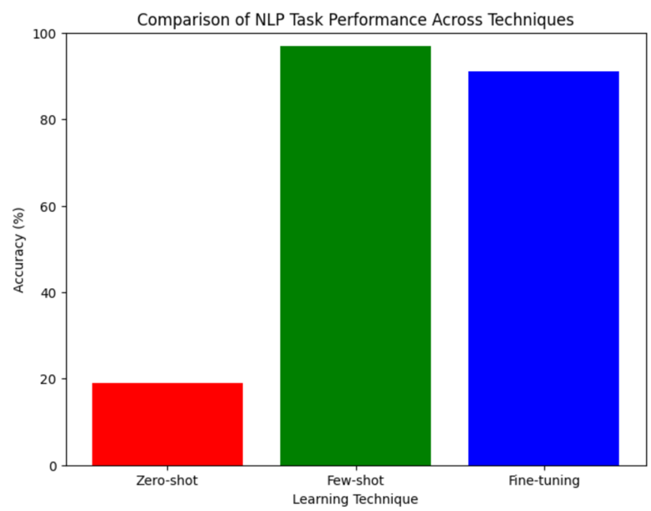

Fig.1. *"For better comprehensive understanding of the differences between zero-shot learning, few-shot learning, and fine-tuning in natural language processing or NLP related works, external content from industry case studies has been referenced. This approach allows for the contextualization of performance benchmarks and enhances the empirical foundation of the study. The use of case studies, like the one depicted by Labelbox, an AI and Data Science platform, validates the comparative analysis by showing the performance of each technique in real-world applications. Furthermore, the analysis of results from external experiments strengthens the generalizability of findings and supports the broader applications and uses of the proposed methods in NLP tasks."*

III. Proposed Methodology

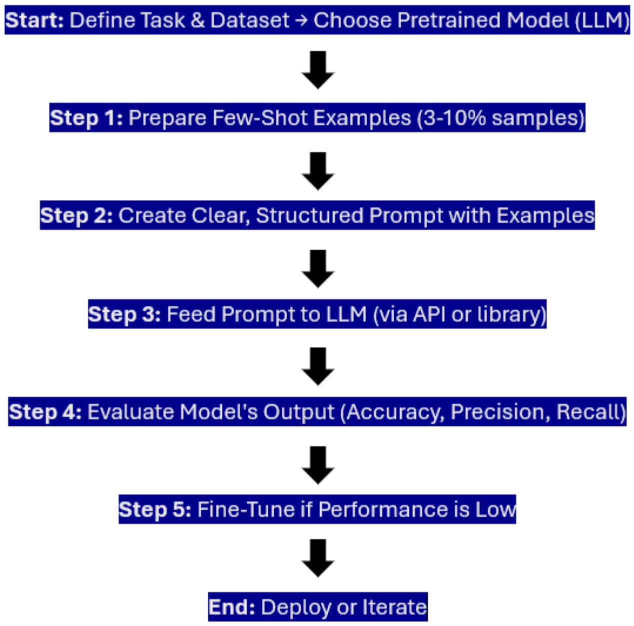

Fig.2. A flowchart of working of few shot learning

A. . Data Collection and Preprocessing

Data preprocessing is an extremely significant step in natural language processing tasks as it affects the material and performance of the model, talking of medical transcription. The text often includes domain-specific terminologies, abbreviations, and medical jargon that must be processed properly to retain their meaning. The following preprocessing steps are applied to the dataset:

- Loading the Data: The first step involves loading the dataset containing medical transcriptions. The mtsamples.csv file consists of transcriptions paired with corresponding medical specialties. We start by cleaning the dataset, followed by removing rows with missing or null values to ensure that the model training is done on clean, high quality data. Additionally, empty strings in the transcription column are refined out to avoid training on irrelevant data.

```
dataset = pd.read_csv('/content/mtsamples.csv')
dataset = dataset.dropna()  # Dropping rows with missing values
dataset = dataset[dataset['transcription'].str.strip() != '']  # Remove empty transcriptions
```

- Tokenization: After cleaning, tokenization is applied. It splits the transcription text into individual words or tokens, which enables the model to work with smaller and relevant units of text. This is a typical approach for handling text in many NLP models, including transformer-based models such as BERT and DistilBERT. For tokenization, we make use of the NLTK word_tokenize function.

```
from nltk.tokenize import word_tokenize
dataset['transcription'] = dataset['transcription'].apply(word_tokenize)
```

- Stemming: Medical terms often has many variations due to diverse inflections or grammatical forms. To reduce this complexity, we use stemming using the Snowball Stemmer. Stemming lessens the words to their simpler form, helping to normalize the data. For example, words like “running,” “runner,” and “ran” would all be cut short to “run.”

```
from nltk.stem.snowball import SnowballStemmer
sb = SnowballStemmer('english')
def stem_it(text):
  return [sb.stem(word) for word in text]
dataset['transcription'] = dataset['transcription'].apply(stem_it)
```

- Stop word Removal: In a lot of NLP tasks, it is very common to remove words that don’t contribute significantly to the context of the text. These are known as stop words for e.g., “and,” “the,” “is”. In such cases, we use a very simple rule to remove the words with lesser than 3 characters, as they are usually stop words or noise.

```
def stopword_removal(text):
  return [word for word in text if len(word) > 2]
dataset['transcription'] = dataset['transcription'].apply(stopword_removal)
```

- Text Formatting: Finally, the tokens are connected back into strings to ensure they can be used as input into the model. Furthermore, we restrict the transcription text to 512 characters to adhere to the input size constraints of transformer models like DistilBERT.

```
dataset['transcription'] = dataset['transcription'].apply(lambda x: x[:512])  # Truncate to 512 characters
```

These preprocessing steps ensures that we have a clean, consistent, and appropriately formatted data for model training.

B. Few-Shot Sampling

A prominent challenge in medical transcription tasks is the lack of labelled data, especially for some specific medical specialties[1]. Gathering extensive quality and quantity of labelled data is both time-taking and expensive, as medical professionals must annotate the data. To overcome this, we employ a few-shot learning technique , which allows the model to learn efficiently from nominal labelled instances. Few-shot learning is particularly well-suited to situations where large datasets are not available.

- Few-Shot Sampling: To simulate few-shot learning, we firstly select a small, balanced subset of the dataset. A fixed amount of samples per class (e.g., 10 samples) are drawn from each medical specialty. This ensures that the model receives a variety of examples from different classes while keeping the dataset manageable.

```
target_samples_per_class = 10
few_shot_samples = dataset.groupby('medical_specialty').apply(
    lambda x: x.sample(n=min(len(x), target_samples_per_class), random_state=42)
).reset_index(drop=True)
```

- Tokenization for DistilBERT: After selecting the few-shot samples, the transcription text is tokenized with the DistilBERT tokenizer. DistilBERT is a smaller and speedy version of the BERT model, making it well-complying with limited computational resources. Tokenization converts the raw text into numerical representations, which are required for feeding the text into the neural network model.

```
from transformers import DistilBertTokenizerFast
tokenizer = DistilBertTokenizerFast.from_pretrained('distilbert-base-uncased')
encodings = tokenizer(list(few_shot_samples['transcription']), truncation=True, padding=True, max_length=512)
```

- Label Encoding: The mapping of medical specialties and numerical labels is done. This is necessary for any classification task, as machine learning models require numerical inputs for learning. We use a simple dictionary for mapping each medical specialty to a unique integer.

```
label_mapping = {label: idx for idx, label in enumerate(few_shot_samples['medical_specialty'].unique())}
few_shot_samples['label'] = few_shot_samples['medical_specialty'].map(label_mapping)
```

- Train-Test Split: The dataset is then split for training and testing sets with a ratio of 80-20 . The training part is used to fine-tune the model, while the testing set performs evaluation of its generalization ability on unseen data.

```
from sklearn.model_selection import train_test_split
train_texts, test_texts, train_labels, test_labels = train_test_split(
    encodings['input_ids'], few_shot_samples['label'], test_size=0.2, random_state=42
)
```

C. *Model Architecture*

To further carry forward the medical transcription task, we use DistilBERT for sequence classification. DistilBERT is a refined class of BERT which is a powerful transformer based model that has achieved great outcomes on a widespread range of NLP tasks. DistilBERT is chosen because it is lighter and faster than BERT, making it more practical for tasks with limited resources or nominal datasets as per [22].

- Model Initialization: We initialize the DistilBERT model for sequence classification, and then specify the number of output labels equalizes to the quantity of unique medical specialties in dataset.

```
from transformers import DistilBertForSequenceClassification
model = DistilBertForSequenceClassification.from_pretrained('distilbert-base-uncased', num_labels=len(label_mapping)).to(device)
```

- Training Setup: Training hyperparameters like batch size, learning rate and times of epochs, are configured. The rate of learning is set to 2e-5, that is usually used in fine tuning transformer models. The no. of epochs is set to 3, which is generally used for fine tuning on a small dataset.

```
from transformers import TrainingArguments
training_args = TrainingArguments(
    output_dir='./results',
    evaluation_strategy="epoch",  # Evaluate at the end of each epoch
    learning_rate=2e-5,
    per_device_train_batch_size=8,
    per_device_eval_batch_size=8,
    num_train_epochs=3,
    weight_decay=0.01,
    report_to="none"  # Disables logging to WandB
)
```

- Trainer Initialization: The Trainer version from Hugging Face's Transformers library is needed to handle training and evaluation. This makes the process of model training easier by abstracting away much of the boilerplate code.

```
from transformers import Trainer
trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=train_dataset,
    eval_dataset=test_dataset,
)
```

- Model Training: The model is trained using the few-shot samples selected. The trainer takes care of the training loop, and adjusts the weights of model to minimize the loss function and improvises its performance on the training data.

trainer.train()

D. Model Evaluation

Post training the model, evaluation of its working using several metrics, including accuracy, precision, F1 score and recall is done. These metrics provide us a thorough understanding of how good model performs, especially in terms of classifying medical transcriptions to their respective specialties.

- Prediction and Evaluation: The model gives predictions on the test dataset

PSEUDOCODE

1. **Input**: Dataset containing medical_transcriptions and medical_specialties labels.
2. **Data Preprocessing**:
   - Clean dataset by removing rows with missing values.
   - Tokenize transcription texts into tokens using word_tokenize.
   - Apply stemming (e.g., using Snowball Stemmer) to reduce words to their root form.
   - Remove stop words (e.g., words with fewer than 3 characters).
   - Truncate transcription texts to 512 characters to adhere to model input constraints.
3. **Few-Shot Sampling**:
   - For each medical specialty, select a small number (e.g., 10) of samples for training.
   - Create a balanced dataset with N samples from each class.
4. **Model Input**:
   - Tokenize transcriptions using a pre-trained model tokenizer (e.g., DistilBERT).
   - Convert medical specialties into numerical labels.
5. **Model Architecture**:
   - Initialize transformer model (e.g., DistilBERT) for sequence classification with num_labels equal to the number of medical specialties.
6. **Training**:
   - Define loss function: criterion = LabelSmoothingCrossEntropy().
   - Define optimizer: optimizer = optim.AdamW(model.parameters(), lr=0.001).
   - Define learning rate scheduler: scheduler = optim.lr_scheduler.StepLR(optimizer, step_size=3, gamma=0.97).
   - Train model using few-shot training set for num_epochs epochs.
7. **Evaluation**:
   - After training, evaluate the model on the test set using evaluate_model(model, test_loader, criterion).
   - Compute metrics like test_loss and test_accuracy.
8. **Output**:
   - Predict class (e.g., "real" or "spoof") for each transcription using the trained model.
   - Output predicted class label for new transcription input.

Fig.3. Pseudocode

IV. Experimental Settings and Results

A. Dataset

This study tested a dataset of medical transcriptions designated by major to evaluate the strength of few-shot learning in acknowledging classification challenges alongside limited data. The dataset was cautiously processed to make sure of consistency and condition, removing noise and unnecessary intel. Each transcription was associated with a medical major quality, forming the foundation for the classification job.

The examination focused on a scenario where only 10 samples for every class were taken for training. This structure shows real-world conditions where attaining huge amount of annotated data is bold, particularly in specialized spheres such as medical transcription. The dataset was pre-processed using tokenization, stemming, as well as truncation to arrange it for input into a transformer-based model, making sure it may effectively handle the task in spite of the limited count of examples.

B. Model Performance

The working of model was checked in context of its certainty on the classification job. The results showed that the model attained a 92.50% accuracy when trained with just 10 samples every class. This extraordinary result underscores the potential of the model to derive effectively, even with basic training data.

Few-shot learning methodologies, connected with a transformer-based model, performed a pivotal role in obtaining this result. By leveraging pre-trained depiction and calibrating on a small subset of the dataset, the model might catch key patterns and distinctions in the data. This degree of accuracy is evident, given the latent complexity and changeability of medical transcription data, which frequently contains specific domain terminology and differing styles.

The experiment highlighted that the model successfully balanced learning from limited data while avoiding overfitting, as evidenced by its robust performance on the validation set. The result shows the potential of transformer-based architectures for tasks where data is scarce, particularly in specialized fields such as healthcare.

C. Conclusion

The ability of the model to gather 92.50% precision with only 10 samples for every class authenticate the success of few shot learning in addressing data-scarcity problems. This result focuses the ability of model to conclude well, making it applicable for stationing in real-world scenarios where gathering and annotating large datasets is generally not feasible. These findings show that transformer-based architectures, united with few-shot learning, can give us scalable and efficient explanation for medical transcription categorization tasks.

Forthcoming work could delve into evolving this procedure to include multi-label classification or administer more impure medical transcription classes, further improving its usage in wellness applications.

V. Test Findings

The test phase of the experiment was designed for the evaluation of the ability of generalization of the framework on unseen data after being trained with a few-shot learning approach using only 10 samples per class. The outcomes from the testing phase are as follows:

A. Accuracy Performance

The model attained an accuracy of **92.50%** in test set. This result demonstrates the model's powerful classification potential medical transcriptions into their respective specialties, even when trained on a minimum number of examples. The high accuracy indicates that the model effectively captured the underlying structures and relations within the medical transcription data.

B. *Impact of Few-Shot Learning*

The few-shot learning methodology proved to be efficient in addressing the challenges posed by data scarcity. Despite being trained with limited data per class, the model showcased robust performance on the test set. This emphasizes the potential of pre trained transformer frameworks to generalize well when fine-tuned with domain-specific data, even when we have low-resource settings.

C. *Observations on Generalization*

- The model exhibited a constant ability to correctly predict the medical specialty for most test samples, reflecting its strength to generalize patterns learned during training.
- Minor misclassifications were observed in cases where the transcriptions contained overlapping terminologies or obstruse content, a common challenge in medical datasets.
- These findings propose that the model's performance could be further enhanced with additional contextual information or slightly increasing the sample sizes.

D. *Real-World Applicability*

The results validate the preparation of model for real world applications, mostly when labelled medical data is minimal, expensive or difficult to obtain. With its high accuracy, the model can be effectively deployed for automated medical transcription tasks, which will in turn reduce efforts of humans and improving efficiency in healthcare documentation workflows.

```
Accuracy: 0.9250
Precision: 0.9302
Recall: 0.9250
F1 Score: 0.9265
```

Fig.4. Results obtained

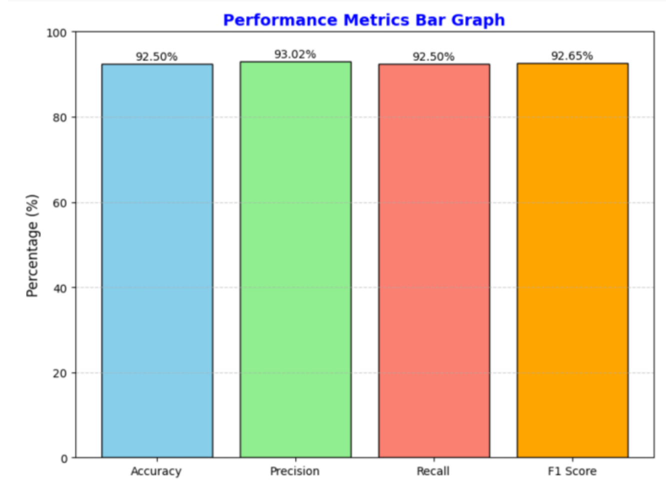

Fig.5. Bar graph of performance

VI. CONCLUSION

This study explored the applications of few-shot learning to improve the classification of medical transcriptions into their respective specialties. By leveraging transformer-based architectures and cutting-edge data preprocessing techniques, the model demonstrated its ability to achieve an impressive 92.50% accuracy with only 10 samples per class. This result highlights the capacity of few-shot learning to address challenges related to data scarcity in specialized fields such as healthcare.

The methodology employed in this research, including proper data cleaning, tokenization, and leveraging pre-trained models, proved efficient in extracting meaningful patterns from limited labelled data. The results emphasize the potential of transformer-based models to generalize well, even in low-resource settings, making them appropriate for real-world applications in medical transcription.

The experimental findings emphasizes the practicality of few-shot learning for scenarios where large-scale labelled datasets are difficult or expensive to obtain. This approach paves the way for efficient and scalable solutions that can be applied in healthcare by automating and streamlining transcription tasks. Such automation not only increases productivity but also reduces human error in crucial documentation processes.

In conclusion, the study validates few-shot learning as an immensely capable tool for solving classification problems in data-constrained environments. Future work could involve extending this approach to multi-label classification, moreover fine-tuning models for domain-specific terminology. Additionally, incorporating external knowledge sources or unsupervised learning methods could also enhance the model's performance and applicability. This research marks a step forward in using innovative machine learning techniques to address critical challenges in the healthcare domain.

REFERENCES

[1] Y. Ge, Y. Guo, S. Das, M. A. Al-Garadi, and A. Sarker, “Few-shot learning for medical text: A review of advances, trends, and opportunities,” Journal of Biomedical, 2023. [Online]. Available: Elsevier.

[2] M. Moor, Q. Huang, S. Wu, M. Yasunaga, et al., “Med-Flamingo: A multimodal medical few-shot learner,” Proceedings of Machine Learning for Health (MLR), 2023. [Online]. Available: mlr.press.

[3] M. Badawi, M. Abushanab, S. Bhat, and A. Maier, “Review of Zero-Shot and Few-Shot AI Algorithms in The Medical Domain,” arXiv preprint arXiv:, 2024. [Online]. Available: arxiv.org

[4] J. Kotia, A. Kotwal, R. Bharti, and R. Mangrulkar, “Few-shot learning for medical imaging,” in Machine Learning Algorithms for Medical Imaging, Springer, 2021.

[5] R . Feng, X. Zheng, T. Gao, J. Chen, et al., “Interactive few-shot learning: Limited supervision, better medical image segmentation,” IEEE Transactions on Medical Imaging, 2021. [Online]. Available: ieee.org.

[6] Y. Ge, Y. Guo, Y. C. Yang, M. A. Al-Garadi, et al., “Few-shot learning for medical text: A systematic review,” arXiv preprint arXiv:, 2022. [Online]. Available: arxiv.org.

[7] S. Ling, A. Pastor, J. Li, Z. Che, et al., “Few-shot pill recognition,” in Proceedings of the IEEE CVF, 2020. [Online]. Available: thecvf.com.

[8] C. Ouyang, C. Biffi, C. Chen, et al., “Self-supervised learning for few-shot medical image segmentation,” IEEE Transactions on Medical Imaging, 2022. [Online]. Available: ieee.org.

[9] A. R. Feyjie, R. Azad, M. Pedersoli, C. Kauffman, et al., “Semi-supervised few-shot learning for medical image segmentation,” arXiv preprint arXiv:, 2020. [Online]. Available: arxiv.org.

[10] Q. Quan, Q. Yao, J. Li, and S. K. Zhou, “Which images to label for few-shot medical landmark detection?” in Proceedings of the IEEE CVF, 2022. [Online]. Available: thecvf.com.

[11] Y. Wang, S. Wang, Y. Li, and D. Dou, “Recognizing medical search query intent by few-shot learning,” in Proceedings of the 45th International ACM SIGIR, 2022. [Online]. Available: github.io

[12] R. Singh, V. Bharti, V. Purohit, A. Kumar, and A. K. Singh, “MetaMed: Few-shot medical image classification using gradient-based meta-learning,” Pattern Recognition, 2021. [Online]. Available: Elsevier.

[13] F. Shakeri, Y. Huang, J. Silva-Rodríguez, H. Bahig, et al., “Few-Shot Adaptation of Medical Vision-Language Models,” in Conference on Medical Vision-Language Models, Springer, 2024. [Online]. Available: arxiv.org.

[14] A. Cai, W. Hu, and J. Zheng, “Few-shot learning for medical image classification,” in International Conference on Artificial Neural Networks, Springer, 2020. [Online]. Available: unsw.edu.au.

[15] Snell, J., Swersky, K., Zemel, R.S.: Prototypical networks for few-shot learning. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017, pp. 4077–4087 (2017)

[16] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H.S., Hospedales, T.M.: Learning to compare: relation network for few-shot learning. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018, pp. 1199–1208. IEEE Computer Society (2018)

[17] V. Joshi, M.E. Peters, M. Hopkins, Extending a Parser to Distant Domains Using a Few Dozen Partially Annotated Examples, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 1190–1199, [Online]. Available:.Google Scholar

[18] Kaiser L., Nachum O., Roy A., Bengio S .Learning to remember rare events International Conference on Learning Representations (2017) [Online]. Available: https://arxiv.org/abs/1703.03129 Google Scholar

[19] Yu M., Guo X., Yi J., Chang S., Potdar S., Cheng Y., Tesauro G., Wang H., Zhou B. Diverse few-shot text classification with multiple metrics.Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana (2018), pp. 1206-1215, 10.18653/v1/N18-1109[Online]. Available.

[20] Rios A., Kavuluru R.Few-shot and zero-shot multi-label learning for structured label spaces.Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing, Vol. 2018, NIH Public Access (2018), p. 3132[Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6375489/ [CrossrefView in ScopusGoogle Scholar]

[21] Z. Hu, X. Li, C. Tu, Z. Liu, M. Sun, Few-shot charge prediction with discriminative legal attributes, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 487–498, [Online]. Available: Google Scholar

[22] I. Chalkidis, M. Fergadiotis, S. Kotitsas, P. Malakasiotis, N. Aletras, I. Androutsopoulos, An Empirical Study on Large-Scale Multi-Label Text Classification Including Few and Zero-Shot Labels, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2020, pp. 7503–7515, eprint. Google Scholar