8-2023

Deep Learning Models For Biomedical Data Analysis

Lucy Nwosu

DEEP LEARNING MODELS FOR BIOMEDICAL DATA ANALYSIS

A Dissertation

By

LUCY NWOSU

Submitted to the Office of Graduate Studies of
Prairie View A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2023

Major Subject: Electrical Engineering

DEEP LEARNING MODELS FOR BIOMEDICAL DATA ANALYSIS

A Dissertation

By

LUCY NWOSU

Submitted to the Office of Graduate Studies of
Prairie View A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Dr. Xishuang Dong Committee Chair	Dr. Lijun Qian Committee Member
Dr. Xiangfang Li Committee Member	Dr. Seungchan Kim Committee Member
Dr. Noushin Ghaffarti Committee Member	Dr. Annamalai Annamalai Interim Head of Department
Dr. Pamela Obiomon Dean, College of Engineering	Dr. Tyrone Tanner Dean, Office of Graduate Studies

August 2023

Major Subject: Electrical Engineering

ABSTRACT

Deep Learning Models for Biomedical Data Analysis

(August 2023)

Lucy Nwosu

Prairie View A&M Universty

Chair of Advisory Committee: Dr. Xishuang Dong

The field of biomedical data analysis is a vibrant area of research dedicated to extracting valuable insights from a wide range of biomedical data sources, including biomedical images and genomics data. The emergence of deep learning, an artificial intelligence approach, presents significant prospects for enhancing biomedical data analysis and knowledge discovery. This dissertation focused on exploring innovative deep learning methods for biomedical image processing and gene data analysis.

During the COVID-19 pandemic, biomedical imaging data, including CT scans and chest x-rays, played a pivotal role in identifying COVID-19 cases by categorizing patient chest x-ray outcomes as COVID-19-positive or negative. While supervised deep learning methods have effectively recognized COVID-19 patterns in chest x-ray datasets, the availability of annotated training data remains limited. To address this challenge, the thesis introduced a semi-supervised deep learning model named ssResNet, built upon the Residual Neural Network (ResNet) architecture. The model combines supervised and unsupervised paths, incorporating a weighted supervised loss function to manage data imbalance. The strategies to diminish prediction uncertainty in deep learning models for critical applications like medical image processing is explore. It achieves this through an

ensemble deep learning model, integrating bagging deep learning and model calibration techniques. This ensemble model not only boosts biomedical image segmentation accuracy but also reduces prediction uncertainty, as validated on a comprehensive chest x-ray image segmentation dataset.

Furthermore, the thesis introduced an ensemble model integrating "Proformer" and ensemble learning methodologies. This model constructs multiple independent "Proformers" for predicting gene expression, their predictions are combined through weighted averaging to generate final predictions. Experimental outcomes underscore the efficacy of this ensemble model in enhancing prediction performance across various metrics.

In conclusion, this dissertation advances biomedical data analysis by harnessing the potential of deep learning techniques. It devises innovative approaches for processing biomedical images and gene data. By leveraging deep learning's capabilities, this work paves the way for further progress in biomedical data analytics and its applications within clinical contexts.

Index Terms-biomedical data analysis, COVID-19, deep learning, ensemble learning, gene data analytics, medical image segmentation, prediction uncertainty, Proformer, Residual Neural Network (ResNet), semi-supervised learning.

DEDICATION

The Almighty God
My husband, Awa Okoro Nwosu

ACKNOWLEDGEMENT

I express my heartfelt gratitude to my supervisor, Dr. Xishuang Dong, for his invaluable guidance, unwavering support, and immense patience throughout the entire duration of my research. His profound knowledge and extensive experience have been instrumental in shaping my academic pursuits and personal growth. Whenever I encountered obstacles or sought advice, Dr. Dong was readily available and skillfully steered me in the right direction.

I extend special thanks to Dr. Seungchan Kim for providing me with invaluable resources and research opportunities at The Center for Computational SystemsBiology, Prairie View A&M University. The research team at the center was also been instrumental in offering technical support and fostering a conducive learning environment. I am sincerely grateful!

I would also like to express my appreciation to the members of my research advisory committee, Dr. Seungchan Kim, Dr. Xiangfang Li, Dr. Lijun Qian, and Dr. Noushin Ghaffari. Their valuable advice, unwavering support, and academic and technical assistance have been integral to the success of this study. Without their passionate involvement and input, this dissertation would not have been accomplished. I am grateful to Dr. Pamela Obiomon, the Dean of Roy G. Perry College of Engineering, Dr. Annamalai Annamalai, the interim Head of the Department of Electrical and Computer Engineering, and Dr. Richard Wilkins, the Graduate Coordinator of the Department of Electrical and Computer Engineering, for building an environment that fosters learning and growth. I would also like to extend my thanks to other faculty members and staff at PVAMU for their kindness, teaching, and educational support.

I am indebted to my husband, Awa Okoro Nwosu, for his unwavering support throughout my graduate studies. He has made it a priority to drive me to and from school for the past six years. Special thanks go to my children Michael, Grace, and Glorious for their understanding and support. I am grateful to my parents for

everything they have provided me in life and for instilling in me the value of education and it saddens me deeply that my mother passed away on February 25, 2023, a few months before the completion of this work. Your prayers, love, and support have carried me thus far. You all have made this journey beautiful!

Above all, I humbly acknowledge and express my profound gratitude to the Almighty God for the gift of life and the grace to complete this work.

This research work is supported in part by the Texas A&M Chancellor's Research Initiative (CRI), by the United States Office of the Assistant Secretary of Defense for Researchand Engineering (OASD (R&E)) under agreement number FA8750-15-2-0119, by the United States Army Research Office under agreement number W911NF-16-1-0496, and by a Panther RISE grant. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should notbe interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Office of the Assistant Secretary of Defense for Researchand Engineering (OASD (R&E)), the Army Research Office, or the United States Government.

NOMENCLATURE

AI	Artificial Intelligence
ANN	Artificial Neural Network
CT	Computed Tomography
CXR	Chest x-Ray
CNN	Convolutional Neural Network
DCA	Deep Count Autoencoder
DL	Deep Learning
DNN	Deep Neural Network
DEG	Differentially Expressed Genes
DT	Decision Tree
ECE	Expected Calibration Error
FCN	Fully Convolutional Networks
GANs	Generative Adversarial Networks
GSEA	Gene Set Enrichment Analysis
MCE	Maximum Calibration Error
MRI	Magnetic Resonance Imaging
NB	Naive Bayes
ORA	Over Representation Analysis
PCA	Principal Component Analysis
PLS-DA	Partial Least Square Discriminant Analysis
VIP	Variable Importance in Projection

RF	Random Forest
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing
RNN	Recurrent Neural Network
scRNA-seq	Single-Cell RNA sequencing
SVM	Support Vector Machine

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER 1

1. INTRODUCTION

Biomedical data such as genome, transcriptome, protein sequences and imaging or spectroscopic data, provides information about expression patterns, splicing variants, localization, protein-protein interaction, and pathway networks related to anorganism or set of organisms [6, 7]. Analysis of these data plays an essential role in understanding how living organisms function, their disease mechanisms, and provides information for improved medical treatment. Artificial intelligence (AI) technologiesare increasingly used for biomedical and healthcare informatics research. Due to recent advancements in technologies, large amounts of biological and clinical datahave been generated and collected at an unprecedented scale, speed, and complexity. This research focused on effective disease treatment and increased patient survivalrate by using deep learning algorithms as a tool for early detection and diagnosis of diseases from biomedical data.

1.1 Data Analysis on Biomedical Data

In recent years, the ability to monitor and observe individual patients' health has never been more data intensive. Though this enables extensive analysis of clinical data for diagnosis and research, understanding and interpreting findings can be quite complex. Several statistical methods are employed for biomedical data analysis [8]. Point estimates like the mean, median, or mode are used to describe the centraltendency of a distribution while standard deviation or variances is used to check the spread

---

This dissertation follows the style of IEEE.

distribution. The random sampling of a normal distribution is achieved with T-test, other distributions apply the Mann-Whitney U-test. T-test can be unpaired for instance, to determine if the blood oxygen saturation levels are different in smokers vs. non-smokers [9] or paired when the same group is observed undertwo different conditions, like comparing muscle oxygen saturation levels of the same participants before and after exercising [10]. Deep learning methods are also usedfor biomedical data analysis and methods such as Convolutional Neural Networks (CNNs) [11] mostly improve prediction performance have pushed the boundaries of what was possible. Problems which were assumed to be unsolvable are now being solved with super-human accuracy.

1.2 Artificial Intelligence for Data Analysis

Artificial Intelligence (AI) is a collection of technologies that is used to extract insights and patterns from large sets of data. Deep learning and machine learningare subsets of artificial intelligence systems used to perform complex tasks in a waythat is similar to how humans solve problems.

1.2.1 Machine Learning Model

Machine learning focuses on developing computer programs that can access data and use it to learn for themselves. The machine learning process begins with observations or data. It looks for patterns in data so it can later make inferences based on the examples provided. Its function can be descriptive, meaning that the systemuses data to explain what happened; predictive, meaning the system uses the data to predict what will happen; or prescriptive, meaning the system will use the data to make suggestions about what action to take. There are three different types of machine learning approaches [12]:

- Supervised learning: supervised algorithms apply what has been learned in the past to new data using labeled examples to predict future events [12By analyzing a known training dataset, the learning algorithm produces an inferred function to predict output values. Examples Support Vector Machine (SVM) [13], Decision Tree (DT) [14], and Random Forest.
- Unsupervised learning: unsupervised algorithms are used when the training data is not labeled [15]. It infers a function to describe a hidden structure from unlabeled data. Different types of clustering algorithms [16] like k-means clustering and hierarchical clustering exist.
- Reinforcement learning: reinforcement learning trains machines through trial and error, using feedback from its own actions and experiences to establish a reward system [17, 18]. This method allows machines and software agents to automatically determine the ideal behavior within a specific context to maximize its performance.

In addition to these approaches, *semi-supervised approach* is another known machine learning method. It uses only few labelled data as training data and the rest of the training data are unlabeled [19]. As labelled data is scarce and expensive, researchers are focusing more on semi-supervised approach. In this work, supervised and semi-supervised learning approaches were employed to classify chest images into Covid-19, pneumonia and normal.

1.2.2 Deep Learning Model

The main challenge in the research area of AI is to develop an efficient and effective system that can imitate the human brain. Therefore, researchers have developed and

designed deep learning models that can mimic the human brain and present more meaningful information of the context [20]. Deep learning is a subset of machine learning that learns and improves on its own by examining computer algorithms [12]

It works with an artificial neural network. These neural networks attempt to simulate the behavior of the human brain by learning from large amounts of data. Deep learning algorithms seek to exploit the unknown structure in the input data distribution to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features [21]. This has aided image classification and image segmentation applications in medical research [22]. It can beused to solve any pattern recognition problem and without human intervention. However, advancements in biomedical data acquisition and analysis have employed larger, sophisticated neural networks, allowing computers to observe, learn, and react to complex situations faster than humans, which has resulted in accurate and robust computer aided disease diagnosis [4].

1.3 Data Analysis via AI Techniques

1.3.1 Image Classification

Image classification plays an essential role in computer-aided medical image analysis for diagnosing diseases. It can be applied in the clinical diagnosis of diabetic retinopathy [23], skin disease [24], breast cancer [25] and lung disease such as pneumonia [26]. In pneumonia detection using chest x-ray images, the input image is usually labelled to represent predefined classes of the disease to be detected. To handle the scarcity of data, a deep transfer learning model was used [26] for binary image classification. It used an ensemble of three convolutional neural network models: GoogLeNet, ResNet-18, and DenseNet-12, to classify the images to pneumonia and

healthy. A multi-classification [27] with more than two classes can be employed if the task is to detect the type of pneumonia present on an image such as bacteria pneumonia, viral pneumonia, mycoplasma, and fungal pneumonia. Histopathology images from the biopsy samples of breast cancer which were captured by a microscope employed binary image classification [28] to classifysamples to benign or malignant tumors. A convolutional neural network (CNN) was used to extract important features which are then classified using a fully connected network.

1.3.2 Image Segmentation

Image segmentation is a computer vision tool for medical image analysis. It diagnoses diseases by classifying the pixel values on an image into predefined classes instead of the class assigned to the input image. It localizes the disease by recognizing the region of interests (ROIs). Segmentation can be applied to segmenting organs and structures such as pancreas [29], skin lesions [30], and the heart [31]. Deep learning models such as U-Net and FCNs have shown great performance in the segmentation of medical images [30] though both are associated with parameter redundancy as well as disappearing gradient when depth increases. To solve these issues, a skin lesion segmentation model with a loss function [30], which improves the Jaccard index of skin lesion image segmentation was used to improve performance in diagnosing the types of skin lesions and the boundary between lesions and normal skin. Medical Image Segmentation with Convolutional Neural Networks (MIScnn) [32], a state- of-the-art intuitive end to end API pipeline which provides training, prediction, as well as fully automatic evaluation, is used for multi-class semantic Kidney Tumor Segmentation [32], with 300 CT scans. Results show a powerful prediction model

based on the standard 3D U-Net model.

1.3.3 Gene Expression Data Analysis

Gene expression data analysis involves studying the activity levels of genes in a particular biological context, often measured using techniques like RNA sequencing (RNA-Seq) or microarrays. It focuses on understanding which genes are active or inactive, and to what extent, under different conditions or in different tissues. The output of RNA-seq differential expression analysis is a list of significant differentially expressed genes (DEGs) [11]. Over representation analysis (ORA) [9] and Gene set enrichment analysis (GSEA) [11] methods are used to gain greater biological insight on the differentially expressed genes. The output from the over representation analysis (ORA), tests whether a gene set contains disproportional genes of significantexpression change. The output of the Gene set enrichment analysis (GSEA) is a test of whether genes of a gene set accumulate at the top or bottom of the full gene vector ordered by direction and magnitude of expression change. To visualize and identify gene sets associated with each subtype, plots such as upset plot, dot plot, ridge plot,barplot, and gseaplot are used. It provides information on the gene count and the enrichment score [11]. In the case of DNA sequences, they are composed of four types of deoxyribonucleotides (bases) (A, T, C, G) that contribute to the diversityof DNA molecules [33]. The DNA double helix structure in Fig. 1.2 demonstrates how the bases from one strand specifically bond with complementary bases on the other strand, forming base pairs as the basic units of DNA sequences.

1.4 Deep Learning Based Image Processing

Image processing transforms an image into a digital form and performs certain operations to get some useful information from it. It uses several deep learning models for image classification and image segmentation tasks. The deep learning models used in this research are described here.

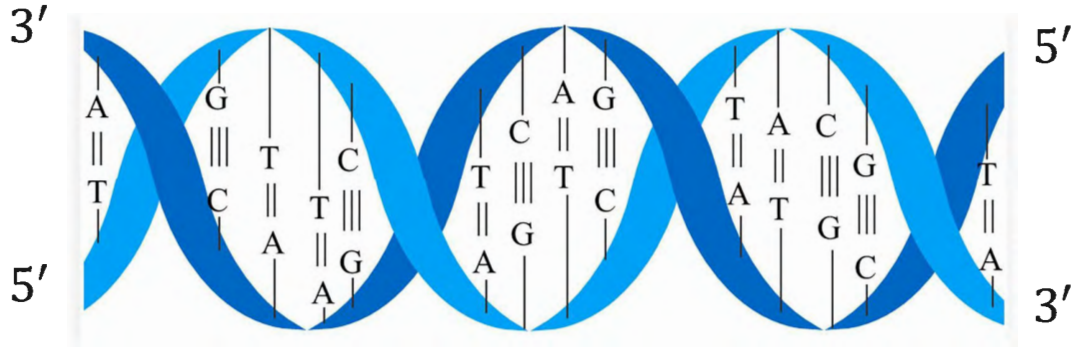


Fig. 1.1. The Double helix of DNA

[33]

1.4.1 Residual Networks (ResNet)

A residual neural network (ResNet) [34, 1] is an artificial neural network (ANN). It aims to solve the problem of the vanishing/exploding gradient. Its architecture introduced the concept called Residual Blocks and uses a technique called skip connections. The skip connection connects activations of a layer to further layers by skipping some layers in between. This forms a residual block which is stacked together to form ResNet. Fig 1.2 shows the relationship between the three building blocks: embedding, mapping and prediction [34].

1.4.2 Fully Convolutional Networks (FCN)

A fully convolutional network (FCN) uses a convolutional neural network to transform image pixels to pixel classes Long et al. [35]. Unlike the CNNs, a fully convolutional network transforms the height and width of intermediate feature maps back to those of the input image. The model first uses a CNN [36] to extract image features, then transforms the number of channels into the number of classes via a convolutional layer, and finally transforms the height and width of the feature maps. iv

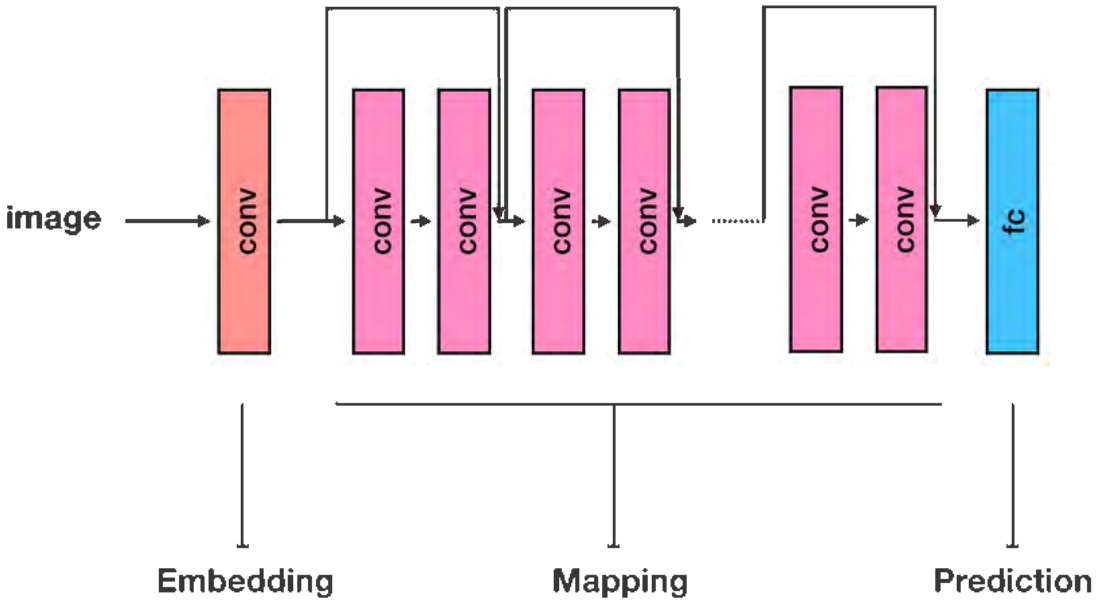


Fig. 1.2. A schematic view of ResNet architecture, decomposed into three blocks: embed-ding, mapping and prediction [1]

1.4.3 U-shaped Encoder-Decoder Network (UNET)

UNET is u-shaped encoder network followed by a decoder network developed for Biomedical Image Segmentation [35, 36, 2]. It requires a discrimination at pixel level but also a mechanism to project the discriminative features learnt at different stagesof the encoder onto the pixel space. The encoder is the first half in the architecture diagram (Fig. 1.3). It uses a pre-trained classification network like VGG or ResNet where convolution blocks are applied followed by a maxpool downsampling to encode the input. The decoder is the second half of the architecture. The goal is to semantically project the discriminative features (lower resolution) learnt by the encoder onto the pixel space (higher resolution) to get a dense classification. The decoder consists of up-sampling and concatenation followed by regular convolution operations [35].represent copied feature maps. The arrows of different colors represent different operations. [2]

1.4.4 MobileNet

MobileNet is a type of convolutional neural network designed for mobile and embedded vision applications. They are based on a streamlined architecture that uses depthwise separable convolutions to build lightweight deep neural networks that can have low latency for mobile and embedded devices. The architecture uses a depthwise separable convolutions to construct lightweight deep convolutional neural networks and provides an efficient model for mobile and embedded vision applications [15]. The depthwise separable convolution filters are composed of depthwise convolution filters and point convolution filters. The depthwise convolution filter performs a single convolution on each input channel, and the point convolution filter combines

the output of depthwise convolution linearly with one-by-one convolutions.

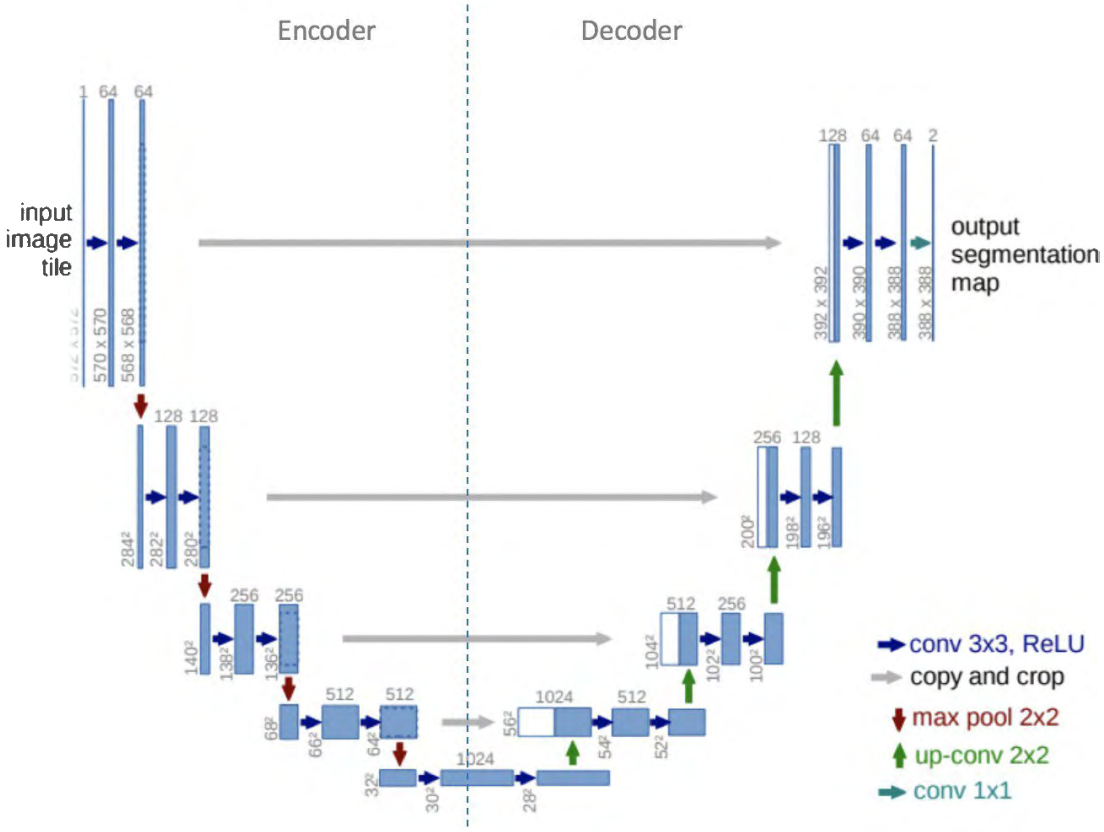


Fig. 1.3. U-net architecture. Blue boxes represent multi-channel feature maps, while boxes

1.4.5 Pyramid Scene Parsing Network (PSPNet)

Pyramid Scene Parsing Network (PSPNet) is a semantic segmentation modelthat utilizes a pyramid parsing module that exploits global context information by different region-based context aggregation [37]. The local and global clues togethermake the final prediction more reliable. The PSPNet architecture considers the global context of the image to predict the local level predictions and hence, gives better performance on benchmark datasets like PASCAL VOC 2012 and cityscapes [38].The model was

designed because FCN based pixel classifiers were not able to capture the context of the whole image.

1.5 Deep Learning Based Gene Expression Prediction

Deep learning has emerged as a powerful technique for analyzing gene expression data, enabling researchers to gain insights into gene function, biological pathways, and disease mechanisms. It provides a powerful framework for analyzing gene expression data, offering improved accuracy in classification, clustering, prediction, and regulatory network inference tasks. These models leverage the expressive capacityof neural networks to extract meaningful representations and capture complex patterns in gene expression profiles. Deep autoencoders are neural network architectures that aim to reconstruct input data by learning a compressed representation in the hidden layers. They have been successfully applied to identify informative featuresand patterns in gene expression data, leading to improved classification and clustering accuracy [39]. Du et al. [40], applied convolutional filters to gene expression profiles. Due to CNNs ability to extract spatial and local dependencies from gene expression data, it can capture important patterns and motifs associated with gene expression patterns, facilitating tasks such as gene classification and bio-marker discovery. Recurrent neural networks (RNNs) have also been employed in modeling the temporal dependencies present in time-series gene expression data. By utilizing the sequential nature of gene expression measurements, [41] used RNNs to capture dynamic patterns and correlations over time, enabling accurate prediction of geneexpression levels and gene regulatory network.

1.5.1 Transformer

Attention-based deep learning models, such as the transformer architecture, can effectively capture long-range dependencies and interactions between genes. It has been applied to gene expression imputation, gene expression prediction, and gene regulatory network reconstruction, yielding competitive performance [42]. This research leverages the transformer model in Fig. 1.4 introduced in the groundbreaking paper, "Attention Is All You Need," [43] for gene expression prediction. Transformer architecture, revolutionized sequence-to-sequence tasks by effectively handling long-range dependencies without recurrence or convolution

- Input: each word in the input sequence is converted into an embedding vector. These embedding vectors are then augmented with positional encoding vectors of the same model length, incorporating positional information into the input representation.
- Encoder: the left half of the Transformer architecture comprises the encoder, consisting of two sub-layers. The first sub-layer employs a multi-head self- attention mechanism, allowing the model to attend to different positions in the input sequence. The second sub-layer is a fully connected feed-forward network with ReLU activation, transforming the input sequence into a continuous representation, which is subsequently passed to the decoder.
- Decoder: situated on the right half of the architecture, the decoder takes itsown predicted output sequence at each time-step as input. Similar to the encoder, positional encoding is applied to augment the decoder input. The decoder block comprises three sub-layers: (a) masked self-attention, where masking prevents the

decoder from attending to future words, (b) decoder attention, allowing the decoder to attend to all positions in the input sequence through the output of the encoder, and (c) a fully connected feed- forward network.

- Output: the output of the decoder undergoes a fully connected layer, followed by a softmax layer, producing predictions for the next values in the output sequence.

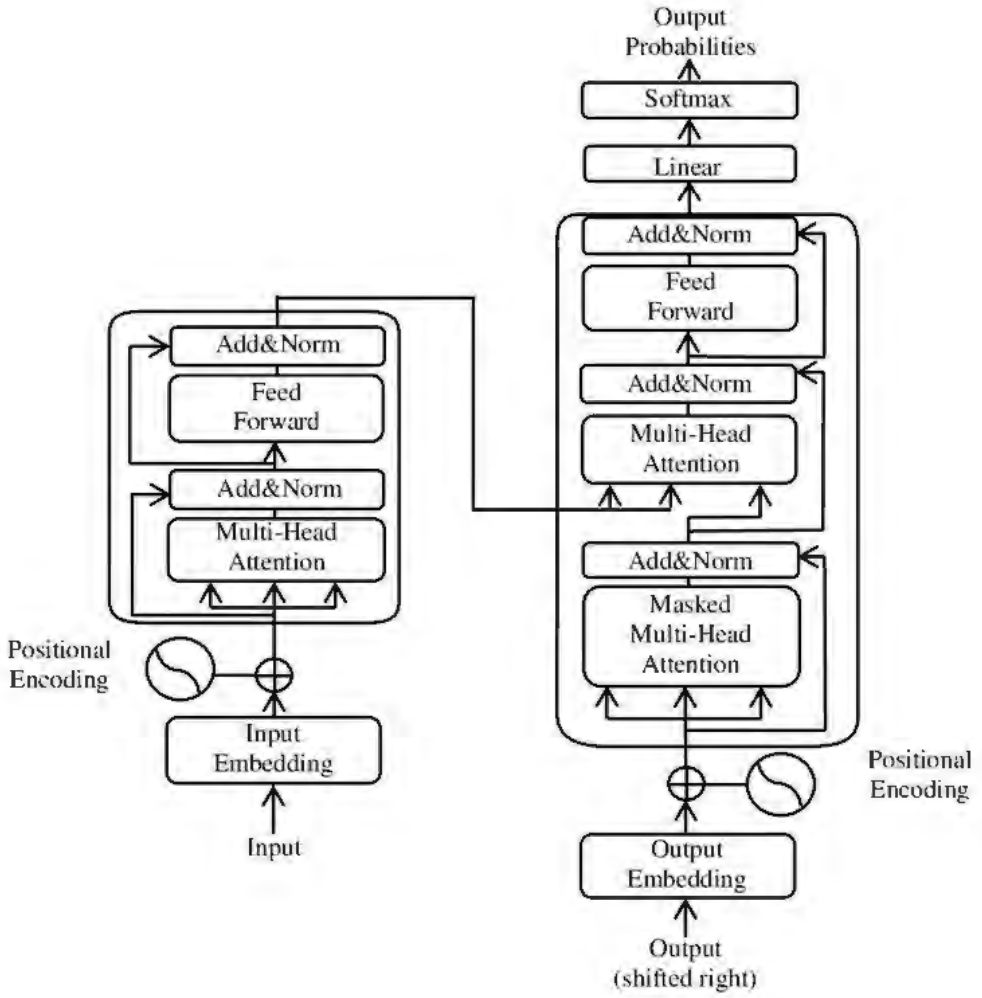


Fig. 1.4. The transformer model architecture[43]

1.5.2 Generative Pre-trained Transformer (GPT) & ChatGPT

The GPT model, developed by OpenAI, is an example of a widely known large language models (LLMs). An LLM can be utilized for gene expression prediction from promoter sequences by leveraging their ability to capture complex patterns and relationships in text data [3]. In this context, the promoter sequence refers to the region of DNA preceding a gene that plays a crucial role in regulating gene expression. To use an LLM for gene expression prediction from promoter sequences, the model needs to be trained on a large dataset of promoter sequences and their corresponding gene expression levels. The promoter sequences can be represented as text strings,where each nucleotide is encoded as a token (e.g., A, C, G, T) [41]. The architecture of GPT in Fig. 1.5 also includes positional embedding, which enables the model to understand the sequential order of words in the input text. By incorporating positional embedding, GPT can consider the relative positions of words and capture the contextual information in the text. GPT is a "unidirectional" model, meaningit generates text in a sequential manner from left to right. This allows the model to learn patterns.

Specifically, ChatGPT, GPT-based chatbot, can be employed to achieve promoter embedding of the data and to implement few-shot learning techniques [33]. During training, the LLM learns the statistical patterns and associations between promoter sequences and gene expression levels. It can capture important sequence motifs, regulatory elements, and other features that influence gene expression. By understanding the context and relationships within the promoter sequence, the model can make predictions about the corresponding gene expression level.

1.6 Challenges

Though many approaches have been proposed to overcome the challenges associated with biomedical data, more research is still required to address these challenges. These are some challenges encountered in this research while working with chest images for Covid-19 detection.

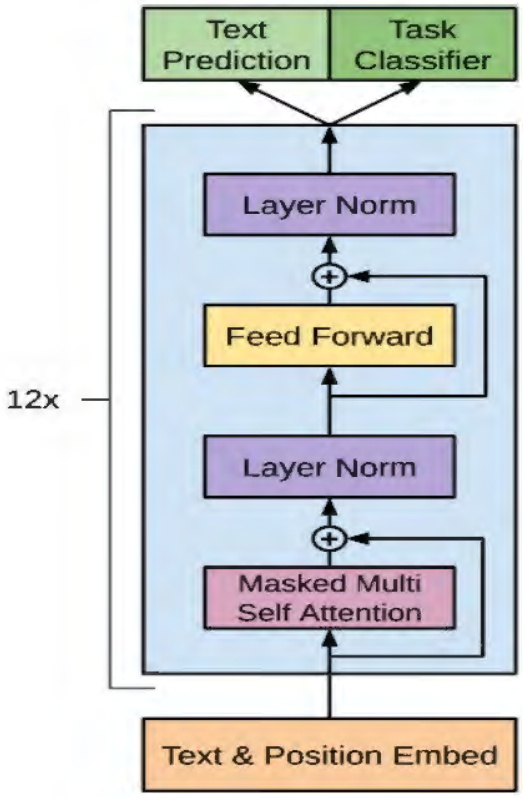


Fig. 1.5. The GPT architecture[43]

1.6.1 Lacking Large Amounts of Labeled Training Data

Accurate and adequate statistical data are essential for research to be effectively carried out for its purpose to be achieved. Biomedical research depends on relevant data from patients, medical centers, hospital, medical practitioners, and other relevant sources. Hence, when the statistical data is adequate and accurate it will aid

biomedical research in providing solutions to the problems of interest which will in the long run assist medical practitioners' efficiency in their jobs. The availabilityof labeled Covid-19 chest images is a big concern in training deep learning modelsto detect Covid-19 patterns as it results in issues such as the inability of models to generalize and under-fitting [44].

1.6.2 Prediction Uncertainty

For safety-critical applications like medical image processing, the prediction uncertainty of DL models is a key evaluation metric on reliability of model predictions, where high prediction uncertainty means low prediction reliability. Prediction uncertainty then arises when uncertainty is characterized through the variance of the prediction rather than through the whole probability distribution [45]. This measureof uncertainty is validated by computing the change of error rate for samples with large prediction regions compared to all samples by benchmarking it on a collection of datasets. For example, for covid-19 applications, applying uncertain predictions to clinical processes would result in disastrous consequences such as missing serious covid cases or delayed treatments. There is need for comprehensive investigation of prediction uncertainty of deep learning models for biomedical imaging task since many deep learning models focus on performance improvement.

1.6.3 Low Performance on Gene Expression Prediction

Understanding the cis-regulatory logic of the genome is of paramount importance as it holds the potential to provide valuable insights into the underlying causes of various diseases. Constructing accurate learning models for gene expression prediction using human data has not achieved satisfactory performance due to

limitations in the diversity of sequences, and the presence of extensive repetitive DNA [41]. Furthermore, biological variability, arising from factors such as cell types, tissue het-erogeneity, and disease states, adds complexity to the data, making it difficult to identify consistent patterns and relationships. This is a big concern in training deep learning models as it results in issues such as training bias, overfitting, over generalization and data-misinterpretation. In addition, the absence of rigorous validation on large datasets limits ~~our~~ understanding of the true potential and reliability of deep learning models for accurately predicting gene expression.

1.7 Problem Statement

1.7.1 Problem Formulation for covid-19 Image Classification and Segmentation

Biomedical image classification on chest images is a supervised learning task which aims to identify predefined classes such as covid-19, normal and pneumonia from chest images. It is considered as either a multi-class image classification or image segmentation problem.

The multi-class classification problem with K chest image classes $=\{s_1, s_2, ..., s_K\}$ where the classes include covid-19, normal and pneumonia class labels and x is the input images, for the image segmentation problem with K pixel classes $S = \{s_1, s_2, ..., s_K\}$, where the pixel classes are the lung region and the background and x are the input images.

$$f(x;\theta) \rightarrow s \qquad (1.1)$$

The aim of this research work was to build a function that could correctly predictthe covid-19 class on chest images. x-ray image dataset Covidx was used to classify chest images into covid-19, normal and pneumonia while the covid-19 chest

x-ray dataset is used to segment and predict each pixel of chest image into lung region and background region. The chest images are fed as input data into deep learning models. Meaningful information was extracted from the data and inferred on the deep learning model for covid-19 image detection. It is very challenging to build a classification model from covid-19 images due to scarcity of labeled data which can result in unreliable models.

1.7.2 Problem Formulation for Gene Expression Prediction

Gene Expression Prediction on millions of randomly generated promoter sequences in yeast is a supervised learning task which aims to predict gene expression level. This is considered a regression problem since the quantitative output of continuous values of gene expressions are handled. The problem statement is as follows:

$$f(x;\theta) \rightarrow k \qquad (1.2)$$

This is a regression problem since the quantitative output of continuous values of gene expression levels $\mathbf{K} = \{k_1, k_2, \ldots, k_k\}$, for each data sample are handled, where θ includes the weight vector w and the bias b. When the promoter sequence x is multiplied elementwise with the weight vector w and summed, the result is combined with the bias term b to obtain the final output of the regression model f(x), which represents the predicted gene expression level based on the given promoter sequenceinput.

1.7.3 Contribution

The major contribution of this research topic was to propose reliable covid-19 detection models from chest images (CT scan, x-ray). The main contributions of the work are based on the two-research work published in this study:

- Contributions from the research Semi-Supervised learning for covid-19 image

Classification Via ResNet.

- Proposed a semi-supervised deep learning model with ResNet through jointly training a supervised ResNet and an unsupervised ResNet. We observed that the proposed model can learn on both unlabeled imagesand labeled images jointly for covid-19 image classification with high performance.
- The proposed model is validated on a large-scale covid-19 image dataset. Experimental results indicate that the proposed model is able to effectively recognize covid-19 images by learning on very few labeled medical images, for example, less than ten percent samples in the training data, which meets the requirement of few available labeled data from the medical do- main for real applications, especially for the cases at the early stage of such a global pandemic.

· Contributions from the research Calibrated Bagging Deep Learning for Image Semantic Segmentation: A Case Study on Covid-19 Chest X-Ray Image

Systematically compare the performance of various state-of-the-art DL models on semantic segmentation on covid-19 CXR data with different evaluation metrics. Moreover, the prediction uncertainty of these DL models were investigated by estimating expected calibration error (ECE) and maximum calibration error (MCE)

- We implemented a novel ensemble deep learning model based on model calibration and bagging deep learning, which is to calibrate bagging deep learning models through weighted summation of predictions generated by individual models. The proposed approach is easily implemented and scalable to various tasks.

- Validate the proposed method on semantic segmentation on a large covid-19 CXR dataset based on different evaluation metrics. Experimental results demonstrated its effectiveness on improving performance and prediction certainty for semantic segmentation.

• Contributions from the research Proformer-based Ensemble Learning for Gene Expression Prediction

- The proposed method is a novel approach that involves an ensemble model consisting of various end-to-end transformer encoders with different architectures. The model predicts the gene expression values of promoter sequences by generating predictions using individual models and combining them by averaging weighted summation.
- I evaluated the proposed model on thousands of randomly generatedpromoter sequences in Yeast using various evaluation metrics. The experimental results demonstrated the effectiveness of the approach in improving the performance of existing state-of-the-art methods through effectivelylearning feature representations of promoter sequences.

1.8 Outline of the Dissertation

This study comprises six chapters and they are framed as follows: the first chapter introduced the theoretical background of the analysis of biomedical data using deep learning methods. This chapter focused on biomedical image processesand posed challenges on the analysis of chest image data for detection of covid-19. Chapter 2 provides literature review on relevant research works. Chapter 3 proposes a novel semi-supervised learning method for covid-19 image classification using ResNet. Chapter 4

introduces a novel method for covid-19 detection basedon a calibrated bagging deep learning method for image semantic segmentation. Chapter 5 introduces a proformer-based ensemble learning for gene expression prediction. Chapter 6 concludes the contribution of this study by highlighting its future work and opportunity.

CHAPTER 2

2. LITERATURE REVIEW

This chapter presents a comprehensive study of recent research work on deep learning models for biomedical data analysis.

2.1Biomedical Data Analysis

There are various types of biomedical data that can be mined to reveal patterns used in making informed medical decisions such as genomics, imaging, proteomics, metabolomics, wearables, and health records. This research focused on the use of genomic and biomedical image data.

2.2Types of Biomedical Data

2.2.1 Genomic Data

The availability of vast amounts of genomic data has revolutionized the understanding of genetics and holds great promise for advancing personalized medicine. Genomic data encompassing DNA sequences, gene expression profiles, epigenetic modifications, and other molecular information and provides a rich resource for studying the complexities of genetic variation and its impact on human health and disease [41]. It provides a foundation for studying gene expression by identifying the genes present in an organism's genome and characterizing their structure and organization [41]. It allows researchers to identify coding sequences, non-coding regions, regulatory elements (such as promoters and enhancers), and other functional elements within the genome. Gene expression can be thought of as the flow of information from the genetic code in DNA to the synthesis of specific molecules that perform essential cellular functions. The process

of gene expression involves several steps [3] as shown in Fig. 2.1.

- Transcription: transcription is the first step in gene expression, where the DNA sequence of a gene is transcribed into a complementary RNA molecule.
- RNA Processing: after transcription, the primary transcript undergoes various processing steps to generate a mature RNA molecule that is ready for translation into a protein or performs a regulatory function.
- mRNA Export: in eukaryotic cells, mature mRNA molecules are transported from the nucleus, where transcription occurs, to the cytoplasm, where translation occurs. This export process involves interactions with specific proteins and nuclear pore complexes, ensuring that only processed and functional mRNA molecules are exported.
- Translation: translation is the process by which the information carried by the mRNA molecule is used to synthesize a protein. It takes place in the cytoplasm and is carried out by ribosomes, complex cellular machinery composed of ribosomal RNA (rRNA) and proteins.
- Post-Translation Modifications: after translation, the newly synthesized polypeptide chain may undergo various modifications to become a functional protein. These modifications include folding into a specific three-dimensional structure, addition of chemical groups (such as phosphorylation or glycosylation), and cleavage of specific regions to generate the final, functional protein.

This research involved predict gene expression levels based on gene expression profiles obtained from high-throughput measurements of the cis-regulatory activity.

of randomly generated promoters in single-cell yeast organisms. The gene expression profiles, and DNA sequences are described as follows:

- Gene Expression Profiles: the analysis of gene expression patterns and levels across a set of genes or an entire genome in a specific biological sample or condition is known as gene expression profiles. They provide a snapshot of the genes that are active or inactive and the extent to which they are ex- pressed in a given context. Gene expression profiles from single-cell sequencing data allows them to explore the transcriptional landscape, identify cell types and states, uncover regulatory dynamics, and gain insights into biological processes or disease mechanisms at the single-cell resolution [41]. By utilizing single-cell sequencing data, researchers can derive gene expression profiles, revealing the intricate patterns and levels of gene expression within a specific biological sample or condition. These profiles enable the exploration of the transcriptional landscape, facilitating the identification of distinct cell types and states. Moreover, they provide valuable insights into regulatory dynamics, allowing researchers to unravel the complexities of biological processes and disease mechanisms at the single-cell resolution [41]. Through the analysis of gene expression profiles, researchers can gain a comprehensive understanding of cellular heterogeneity, uncover novel molecular markers, and shed light on the underlying molecular mechanisms driving cellular behavior.

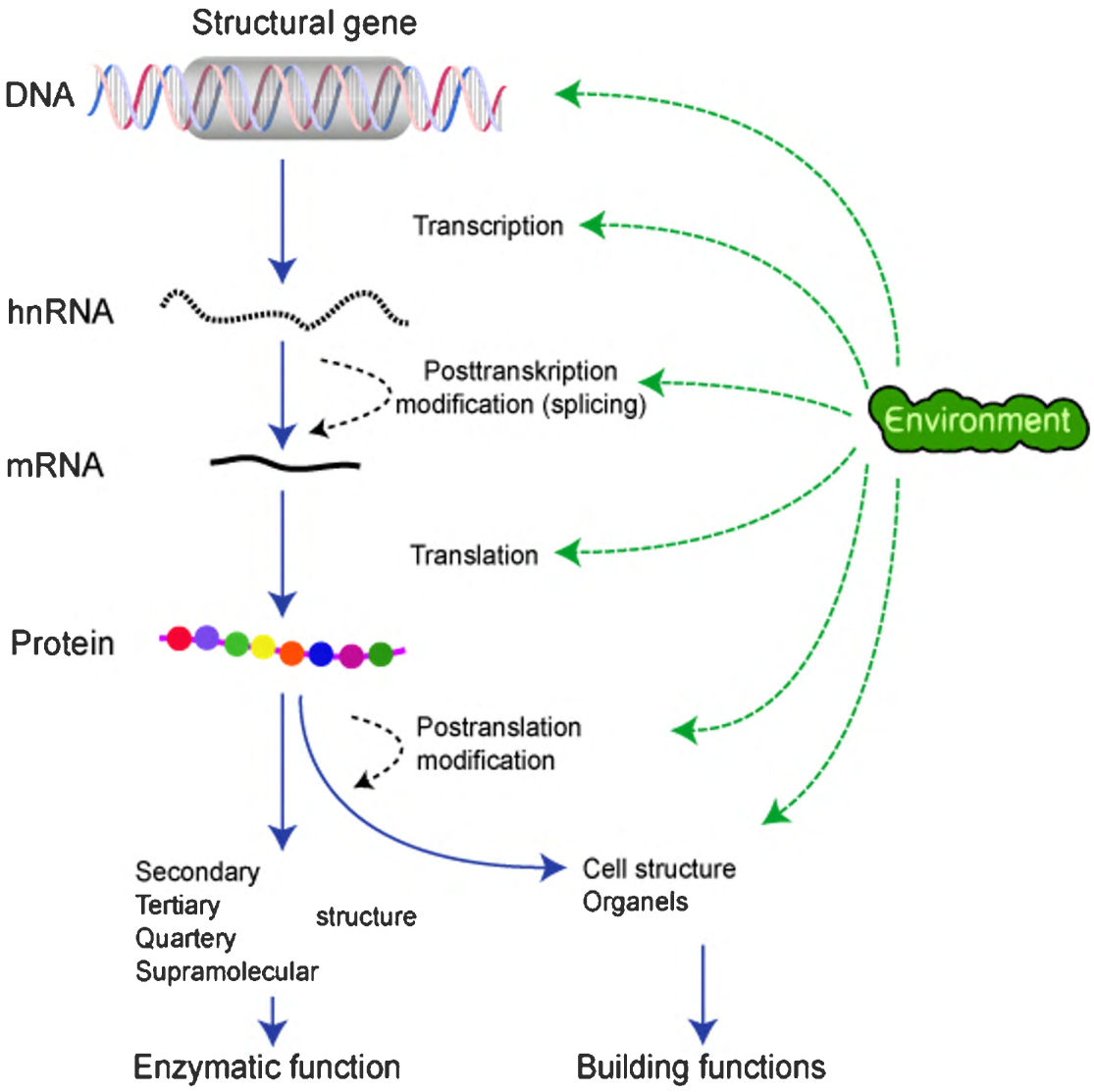


Fig. 2.1. Gene Expression [3]

- DNA sequences: present in the cells of all known organisms, including bacteria, plants, animals, and humans, is the DNA molecule, or deoxyribonucleic acid. It is a long, double-stranded polymer that carries the genetic information in

living organisms. Each nucleotide in DNA is represented by one of four bases: adenine (A), cytosine (C), guanine (G), or thymine (T). For example, a DNA sequence of "GCAAACCAAT" consists of ten nucleotides. DNA sequence refers to the specific arrangement of nucleotide bases in a DNA molecule. It carries the genetic information that determines the characteristics and traits of an organism [41]. DNA sequence refers to the precise order of nucleotide bases in a DNA molecule. It carries the genetic information that determines the characteristics and traits of an organism. It represents the genetic code that carries the instructions for the development, functioning, and inheritanceof all living organisms. A specific region of DNA located upstream of a gene, which contains regulatory elements that control gene expression is known as the promoter sequence. It plays a vital role in initiating the transcription process by binding transcription factors and RNA polymerase, thereby determining when and how strongly a gene is transcribed. Analyzing the promoter sequence helps in understanding gene regulation, identifying transcription factor bindingsites, and elucidating the factors that control gene expression [41]. Promoter sequences can vary between genes and species, and studying their functional elements provides insights into the mechanisms underlying gene regulation and cellular processes.

2.2.2 Biomedical Images

Biomedical imaging is used in biological and health sciences to generate scientifically useful images of various aspects of organisms and biological systems that are not visible to the naked eye [4]. Imaging reveals complex structures and dynamic interactive

processes, located deep inside the body, that are otherwise difficult todetect, Fig. 2.2 shows various biomedical data images such as eye, chest, lung, nerves etc. Some imaging techniques are used to develop images of tissues belowthe skin, while others are used to mark and trace biologically important processes at a molecular level. Clinical image techniques include magnetic resonance imaging (MRI), x-ray computed tomography (CT), ultrasound, and light-based methods -endoscopy and optical coherence tomography (OCT) [21]. Mining and analysis of biomedical images help in diagnosing diseases and conditions, such as chronic obstructive lung disease, pulmonary embolism, lung cancer, brain cancer and covid-19[2, 46]. Below are examples of medical images. x-ray: x-ray imaging is one of the oldest and most commonly used medical imaging techniques. It involves passing a small amount of ionizing radiation through the body to create an image on a film or a digital detector [47]. X- rays are primarily used to visualize bones and the lungs but can also be used toexamine other body parts. They are particularly effective at highlighting dense structures such as bones, allowing for the detection of fractures, dislocations, infections, and certain tumors. x-rays are relatively quick and inexpensive, making them a valuable tool for initial screening and diagnosing a wide range of conditions.

2.3Biomedical Data Analysis and Deep Learning

Biomedical data analysis plays a crucial role in extracting valuable insights from various biomedical datasets, and deep learning models have emerged as powerful tools for this task. These models aid the exploration, interpretation, and extraction of valuable insights from various types of biomedical data [48, 49]. Biomedical data encompasses a wide range of data sources, including biomedical images, genomic

sequences, clinical records, and molecular data. The goal of biomedical data analysis is to uncover patterns, relationships, and meaningful information that can aidin understanding diseases, improving diagnoses, predicting outcomes, and guiding treatment and decisions [50, 51].

Traditional approaches to biomedical data analysis often relied on manual feature engineering, statistical methods, and domain-specific algorithms [52, 49]. However, with the rapid advancement of computational techniques and the availability of large- scale datasets, there has been a paradigm shift towards using machine learning and artificial intelligence methods, particularly deep learning, in biomedical research and clinical applications. Deep learning has revolutionized the field of biomedical dataanalysis by enabling the automatic learning of complex representations directly from raw data. Deep learning models, such as neural networks with multiple layers, have shown remarkable capabilities in capturing intricate patterns and extracting valuable information from biomedical data. These models can effectively handle the high dimensionality and complexity of biomedical datasets, allowing for more ac- curate predictions, improved feature extractions, and enhanced data-driven decision- making [52, 53].

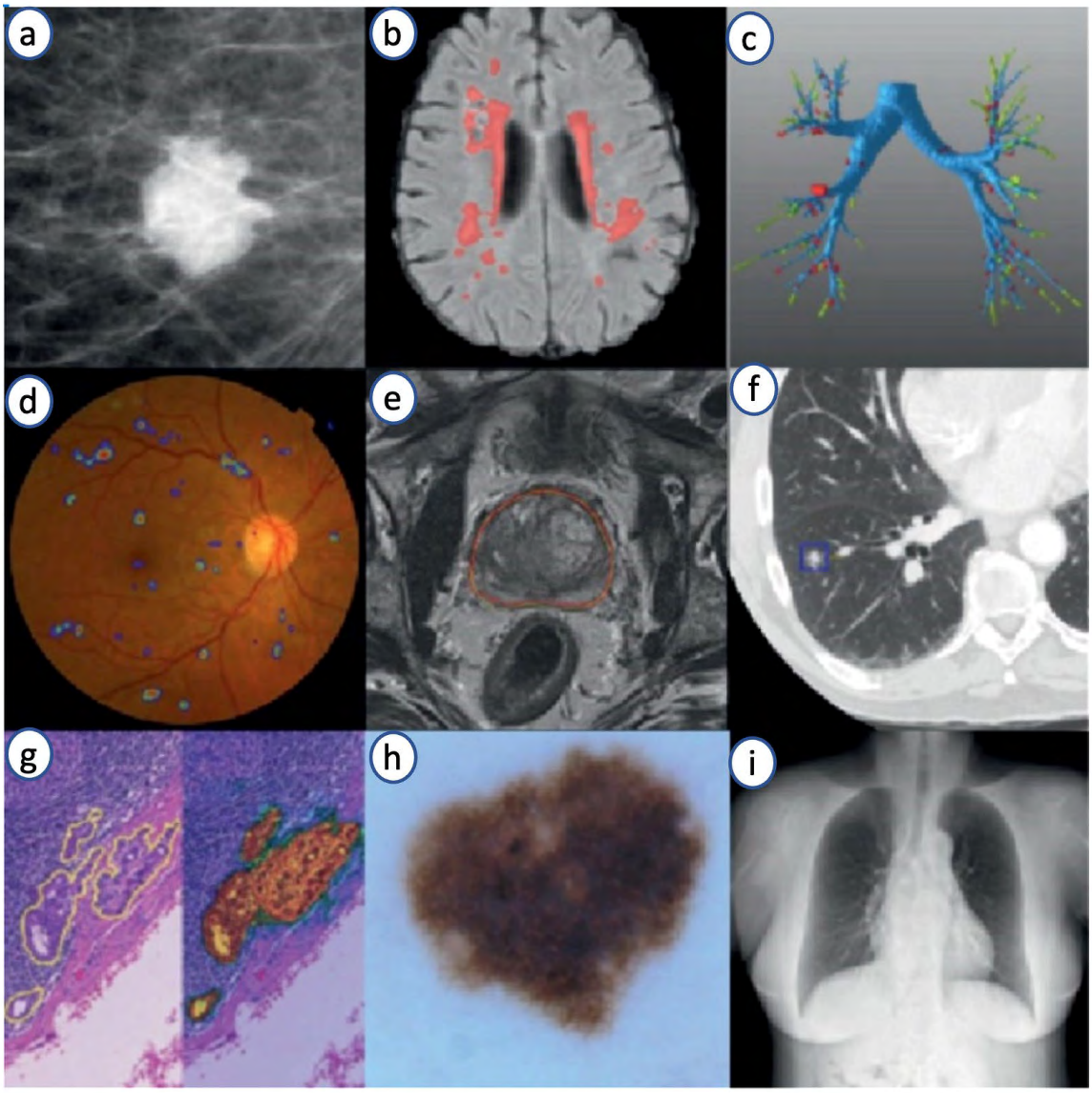


Fig. 2.2. A collection of some biomedical imaging applications in which deep learning has achieved some state-of-the-art results. Images from top-left to bottom-right: (a). Mammographic mass classification, (b). Segmentation of brainlesions, (c). Leak detection in airway tree segmentation, (d). Diabetic retinopathy classification, (e). Prostrate segmentation, (f). Nodule classification. (g). Breast cancer metastases classification, (h). Skin lesion classification, (i). Bone suppression. [4]

2.4Deep Learning Models for Biomedical Image Processing

Deep learning models have emerged as powerful tools for analyzing biomedical images, revolutionizing the field of biomedical image processing. These models leverage the hierarchical representations learned from large-scale datasets to extract meaningful features and facilitate accurate analysis and interpretation of biomedical images. Convolutional Neural Networks (CNNs) [53] have played a pivotal role in advancing biomedical image analysis. CNNs excel in capturing local and global spatial dependencies within images, enabling tasks such as image segmentation, object detection, and disease classification. For instance, U-Net [54], a popular architecture based on CNNs, has demonstrated remarkable success in biomedical image segmentation tasks, such as delineating tumor boundaries in medical scans [53, 55].

Transfer learning has further enhanced the effectiveness of deep learning models for biomedical image processing [55, 56, 57]. By leveraging pre-trained models on large-scale datasets, such as ImageNet, and fine-tuning them on biomedical image data, transfer learning allows for efficient training even with limited labeled biomedical images [56]. This approach has been widely adopted in various biomedical imaging applications, including histopathology image analysis, retinal image analysis, and radiology image interpretation. Beyond traditional 2D imaging, deep learning models have also been applied to3D biomedical image analysis. Recurrent Neural Networks (RNNs) and their variants,such as 3D convolutional neural networks (3D CNNs) [56, 57, 52] and long short- term memory (LSTM) networks [52], have shown promise in handling the temporal and spatial dependencies in volumetric medical images, such as computed tomography (CT) scans and magnetic resonance imaging (MRI)

volumes. These models enable tasks like tumor segmentation, organ tracking, and disease progression monitoring [52]. Generative Adversarial Networks (GANs) [57, 52] have emerged as powerful toolsfor biomedical image synthesis and augmentation. GANs can generate realistic and diverse biomedical images, aiding in data augmentation and addressing the issue oflimited annotated datasets [52, 56, 57, 58]. This facilitates training deep learning models with improved generalization and robustness. GANs have found applicationsin generating synthetic medical images, such as brain MRI scans, retinal fundus images and chest x-rays and computed tomography (CT) images.

Deep learning techniques application to chest images are further described in this research in terms of image localization, segmentation, registration, and classification leading to covid-19 detection. Numerous studies have focused on developing deep learning models specifically for covid-19 image classification. For instance, Wang et al [58] proposed a deep learning framework called COVID-Net, which utilizesa lightweight convolutional neural network (CNN) architecture to classify chest x-ray images into covid-19, pneumonia, or normal cases. Similarly, Apostolopoulos and Mpesiana [59] introduced a deep learning model based on a pre-trained CNN called COVIDX-Net, achieving high accuracy in covid-19 classification. In addition to covid-19 image classification, accurately identifying the regions of infectionin covid-19 is crucial for diagnosis and detailed assessment. Semantic segmentation techniques can help recognize these regions and patterns, allowing for the quantify cation and assessment of covid-19 in chest x-ray or CT images. The regions of interest (ROIs) typically include the lung, lobes, bronchopulmonary segments, and infected regions or

lesions. Deep learning approaches have significantly advanced the field of semantic segmentation in biomedical image analysis [2, 46]. Several segmentation networks have been developed specifically for covid-19, such as the classic U-Net [60, 54, 61], UNet++ [31], and VB-Net [62]. These methods can be classified into two groups: lung-region-oriented methods and lung-lesion-oriented methods. The former focuses on separating lung regions (that is, the whole lung and lung lobes) from other background regions in CT or x-ray images [63, 64]. The latter aims to detect lesions or artifacts within the lung region, such as metal and motion artifacts [65, 66].It is worth noting that segmenting x-ray images is particularly challenging due to the presence of ribs that project onto soft tissues in 2D. While supervised deep learning models outperform other methods in these tasks, they often require a large amount of labeled data for training, which may not be practical in real-world applications. Semi-supervised deep learning has gained attention for its ability to generalize model performance by leveraging both labeled and unlabeled data [67, 68, 69, 70]. In safety-critical applications like medical image processing, autonomous driving, and precipitation forecasting, high accuracy alone is not sufficient. It is equally important to measure the prediction uncertainty and ensure model calibration. Two categories of methods are commonly used to calculate model calibration [71, 72].

Overall, the application of deep learning models in biomedical data analysis, particularly in the context of covid-19, has shown promising results in semantic segmentation for identifying the regions of infection. The use of semi-supervised learning and model calibration techniques further enhances the generalization and reliability of these models in safety-critical applications. While deep learning models have shown

significant advancements in biomedical image processing, challenges remain. Interpreting and explaining the decisions made by deep learning models in the medical domain is a critical concern. Efforts are underway to develop explainable AI techniques that can provide insights into the decision-making process of these models, enhancing their trustworthiness and facilitating clinical adoption These models learn hierarchical representations of image features, enabling precise identification of anatomical structures, early detection ofabnormalities, and accurate diagnosis.

2.5Deep Learning Models for Genomic Data Analysis

Deep learning models have shown great promise in the field of genomic data analysis, enabling researchers to extract meaningful insights and unravel the complexitiesof the genome. By leveraging the power of neural networks and their ability to capture intricate patterns in large-scale genomic datasets, deep learning approaches have revolutionized various aspects of genomic data analysis. It is employed to tackle challenges such as high-dimensional gene expression data, identification of disease-related genetic variations, and prediction of molecular interactions. Recurrent neural networks (RNNs) [41] and transformer-based architectures [73] have shown promise in modeling the sequential and temporal dependencies in genomic sequences, enabling accurate gene expression prediction, classification of disease subtypes, and discovery of biomarkers.

Deep learning models have been applied to variant calling and interpretation. Convolutional neural networks (CNNs) have been utilized to analyze DNA sequences and identify genomic variations, such as single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). By learning informative sequence motifs [74], CNNs can accurately classify genomic variants and aid in the identification of disease-associated

mutations. Deep learning models have been instrumental in deciphering the three-dimensional structure of the genome and understanding its regulatory mechanisms. Convolutional and recurrent neural networks have been employed to analyze chromatin conformation capture data, enabling the prediction of chromatin interactions and identification of regulatory elements [75, 76]. These models have shed lighton gene regulation networks and enhancer-promoter interactions, offering valuable insights into gene expression regulation and cellular processes.

Deep learning models have been successfully employed for gene expression prediction, which plays a crucial role in understanding gene function, biological pathways,and disease mechanisms. Gene expression prediction is an essential task. Gene ex- pression involves using the information stored in genes to create a functional gene product. By connecting the expression of genes of interest to a biological process or phenotype, researchers can gain insights into gene function, biological pathways, and the genes responsible for regulating development, cell behavior, and signaling [75]. The degree of a gene's expression is primarily governed by multiple input signals that are interpreted by the non-coding regulatory DNA sequences known as cis-regulatory logic. These sequences exert control over gene expression intensity by using transcription factors (TFs) that bind to regulatory sequences located throughout the DNA, including promoters that contain a wealth of information related to mRNA levels [75]. Recurrent neural networks (RNNs) and transformer-based architectures have demonstrated their effectiveness in modeling the sequential dependencies present in gene expression data [77, 78, 77]. These models can capture the temporal dynamics of gene expression patterns and provide accurate predictions, facilitating the identification of potential

biomarkers and therapeutic targets.

CHAPTER 3

3. METHODOLOGY

This chapter focuses on the methods for deep learning data analysis. Three methods will be discussed.

3.1 Semi-Supervised Learning for Covid-19 ImageClassification Via ResNet

Coronavirus disease 2019 (COVID-19) outbreak has led to the heavy losses of the world's economy and life. To reduce the spread of covid-19 and the death rate, it is essential to detect the disease at the early stage with effective and timely screening/testing and place covid-19 infected patients in quarantine immediately [79, 80]. Artificial intelligence (AI), an emerging technology for medical imaging processing, has actively contributed to the fight against covid-19 [81]. Compared to the traditional imaging workflow that heavily relies on human interpretation, AI enables more safe, accurate and efficient imaging solutions.

Recent AI-empowered applications in covid-19 detection include the dedicated imaging platform, the lung and infection region segmentation, as well as the clinical assessment and diagnosis [82, 61, 60]. Moreover, commercial products integrate AI to combat covid-19 and demonstrate the capability of the AI technology [82]. All of these examples show the tremendous enthusiasm cast by the public for AI-empowered progress in the medical imaging field, especially during the ongoing covid-19 pan- demic.

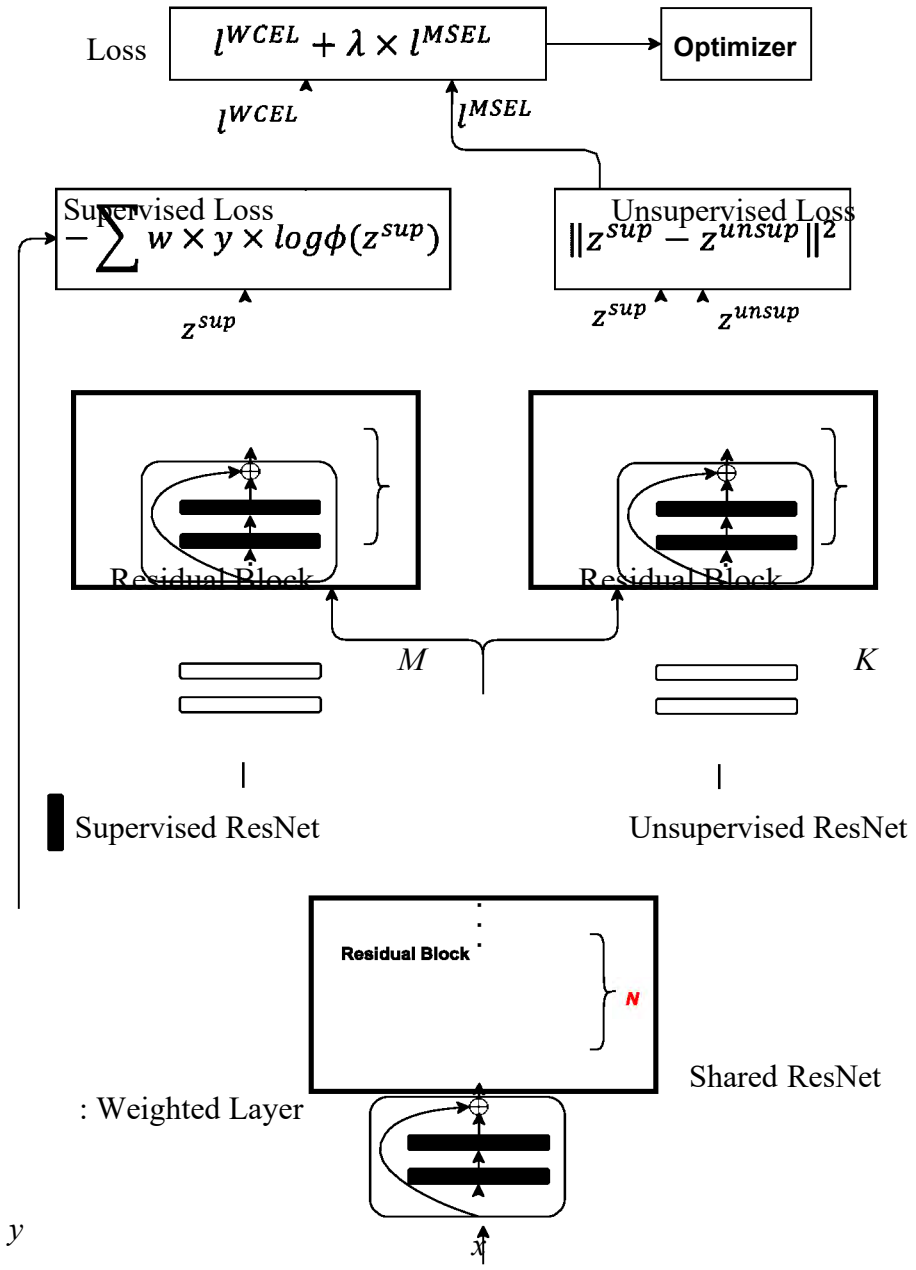


Fig. 3.1. Framework of the proposed semi-supervised learning. Input x is the medical image. Labels such as y are available only for the labeled inputs. SharedResNet will evaluate the input to obtain the low-level representations as inputs to supervised ResNet and unsupervised ResNet, where these three ResNets are builtwith residual blocks and N, M, and K are numbers of residual blocks for these three ResNets. Then z^{sup} and z^{unsup} are outputs from the supervised ResNet and the unsupervised ResNet, respectively. Moreover, z^{sup} and y will be applied to calculate a weighted cross entropy loss l^{WCEL} whereas z^{sup} and z^{unsup} are used to calculate a mean squared error loss l^{MSEL}, where w the weight to different classes of samples. I jointly optimize the combined losses, where λ is the weight for unsupervised loss, ⊕ is the short-cut connection in the residue operation.

Regarding the covid-19 research based on AI, covid-19 image classification becomes more and more attractive, which is to separate covid-19 patients from non-covid-19 subjects using the features extracted from medical images. Specially supervised deep learning such as convolutional neural networks (CNN) has been very popular in this research area. For example, Wang et al. proposed a 2D CNN supervised model to analyze delineated region patches to accomplish classification between covid-19 and typical viral pneumonia [83]. Similarly, Xu et al. utilized candidate infection regions to complete COVID-19 classification via supervised ResNet-18 [84]. In addition, as a powerful deep learning model for medical image analysis, UNet [2] was employed for COVID-19 image classification and segmentation. For example,Zheng et al. employed UNet to obtain lung segmentation and predicted the probability of COVID-19 with 3D CNN on segmentation features [60]. Jin et al. proposeda UNet++ based segmentation model for locating lesions and built a ResNet-50based classification model for COVID-19 diagnosis [63]. Chen et al. implementedCOVID-19 classification with the patterns of segmented lesions extracted by supervised UNet++ [85, 31]. Moreover, they employed a 2D Deeplab model for the lungsegmentation and a 2D ResNet-152 model for lung-mask slice-based identificationof positive COVID-19 cases [86]. Although supervised deep learning presents impressive performance on COVID-19 image classification, it requires a large amountof annotated medical images to train models, which is not practical with respect to limited data resources related to COVID-19, due to huge costs of labeling medical images, and labeling noise [4].

To reduce the efforts on labeling medical images for COVID-19 image

classification, I built a two-path semi-supervised deep learning model that was able to learn on both labeled and unlabeled medical images, based on residual neural net- works (ResNet) [87]. ResNet is an artificial neural network developed by mimicking pyramidal cells in the cerebral cortex. It introduces a so-called "identity shortcut connection" that skips one or more layers since stacking layers should not degrade the network performance. With ResNet, I implemented a two-path semi-supervised learning model that was composed of three components namely, shared ResNet, supervised ResNet, and unsupervised ResNet.

Framework of the proposed model is shown in Fig. 3.1. The right path is composed of a shared ResNet and a supervised ResNet while the left path consists of theshared ResNet and an unsupervised ResNet. All data (labeled and unlabeled data) was evaluated to calculate the unsupervised loss, that is the mean squared error loss (MSEL), while only labeled data was used to calculate the supervised loss, that is the cross-entropy loss (CEL). Specifically, I designed a weighted cross entropy loss (WCEL) that assigns more weight to the COVID-19 class for addressing the data imbalance. Reducing MSEL is to enhance the image representation while decreasing WCEL is to enhance classification performance. The proposed model was validated on a large-scale of x-ray image dataset COVIDx and experimental results demonstrated the proposed model could accomplish covid-19 image classification with promising performance even when trained on the extremely limited amount of labeled x-ray images.The contributions in this study are below.

- A semi-supervised deep learning model with ResNet through jointly training a supervised ResNet and an unsupervised ResNet was proposed. I observed that the

proposed model can learn on both unlabeled images and labeled images jointly for COVID-19 image classification with high performance.

- The proposed model was validated on a large-scale COVID-19 image dataset. Experimental results indicated that the proposed model was able to effectively recognize COVID-19 images by learning on very few labeled medical images, for example, less than ten percent of samples in the training data, which met the requirement of few available labeled data from the medical domain for real applications [4], especially for the cases at the early stage of such global pandemic.

3.1.1 Proposed Methodology

I proposed a semi-supervised ResNet to address the challenge of lacking of labeled data for covid-19 image classification, where the detailed framework is shown in Fig. 3.1. The shared ResNet will generate a new representation z below of inputx.

$$z = f_{pooling}(f_{Resblock_N} \cdots f_{Resblock_1}(x')) \ . \tag{3.1}$$

where

$$x' = f_{conv}(x). \tag{3.2}$$

$$f_{Resblock}(x') = x + f_{conv}(f_{conv}(x')) \ . \tag{3.3}$$

$f_{cov}(\cdot)$ is the convolutional operation. $f_{Resblock}(\cdot)$ is the residual operation [87] and $f_{Resblock_N}$ refers to N sequencing residual operations. $f_{pooling}(\cdot)$ is the pooling operation. The shared ResNet introduced to the proposed model is inspired by deep multi-task learning [99, 100], since different tasks share a low-level feature representation extracted from the input x.

In addition, the reason for learning low-level feature representations instead of directly using x is that the original representation may not have enough expressive power for multiple tasks [101]. With the training data in all tasks, a more powerful representation can be learned for all tasks and this representation will improve performance. As shown in Fig. 3.1, I have two "tasks" in the proposed model, namely, a supervised task and an unsupervised task, which is similar to the framework of deep multi-task learning. Therefore,the shared ResNet is necessary to feed the low-level representations to these twotasks.

The output z from the shared ResNet is evaluated by two ResNets, namely, asupervised ResNet and an unsupervised ResNet. For the supervised ResNet, it is to learn the deep features of labeled samples. The output z^{sup} of the supervised ResNetis given by

$$Z^{sup} = f_{pooling}^{sup}\left(f_{Resblock_M}^{sup}\cdots f_{Resblock_1}^{sup}(Z')\right) \tag{3.4}$$

where

$$Z' = f_{Cconv}^{sup}(Z) \tag{3.5}$$

$$f_{Resblock}^{sup}(Z') = Z' + f_{Cconv}^{sup}(f_{conv}^{sup}(Z')) \tag{3.6}$$

The same operations including the pooling operation f_{conv}^{sup}, the convolution operation $\mathbf{f}^{sup}(\cdot)$ is employed, and M sequencing residual operations $\mathbf{f}_{pool-}^{sup}(\cdot)$. Moreover, I build the unsupervised ResNet to generate another representation of all inputs including labeled data and unlabeled data. This representation $z_{Resbloc}^{unsup}$ is given by

$$Z^{unsup} = f_{pooling}^{unsup}\left(f_{Resblock_M}^{unsup}. \ . \ . f_{Resblock_1}^{unsup}(Z'')\right) \quad (3.7)$$

where

$$Z'' = f_{Cconv}^{unsup}(Z) \quad (3.8)$$

$$f_{Resblock}^{unsup}(Z'') = Z'' + f_{Cconv}^{unsup}(f_{conv}^{unsup}(Z'')) \quad (3.9)$$

Similarly, the pooling operation $f_{pooling}^{unsup}(.)$ is employed, the convolutional operation $\mathbf{f}^{unsup}(\cdot)$, and K sequencing residual operation $\mathbf{f}_{conv}^{unsup}(\cdot)$,are used to build the unsupervised ResNet. Then, those two vectors z^{sup} *and* z^{unsup} *are* utilized to calculate the weighted cross entropy loss (WCEL) and mean squared error loss (MSEL) for supervised and unsupervised paths, respectively. They are given by

$$l^{WCEL} = -\sum w \times y \times \log\phi(z^{sup}). \quad (3.10)$$

$$l^{MSEL} = ||z^{sup} - z^{unsup}||^2 . \quad (3.11)$$

where y is the ground truth of the input and w is corresponding weight. $\phi(\cdot)$ is the softmax activation function. l^{WCEL} is the weighted cross entropy loss to account for the loss of labeled inputs. To enhance classification performance for the minority class (covid-19 class), I assigned more weight to covid-19 class, where during the learning procedure the classifier paid more attentions to covid-19 class so as to reduce the learning bias that was caused by data imbalance.

l^{MSEL} is used to measure the differences between z^{sup} and z^{unsup}. Since training ResNets with dropout regularization and gradient-based optimization is a stochastic process, the two ResNets had different link weights after training. In other words,

there were differences between the two prediction vectors z^{sup} and z^{unsup} that are from these two ResNets (the supervised ResNet and the unsupervised ResNet). These differences can be treated as an error in the classification and thus minimizing this loss is a goal in the proposed model, which is inspired by Π model [70]. Based on these two losses, the total loss is defined by

$$\text{Loss} = l^{WCEL} + \lambda \times l^{MSEL} \tag{3.12}$$

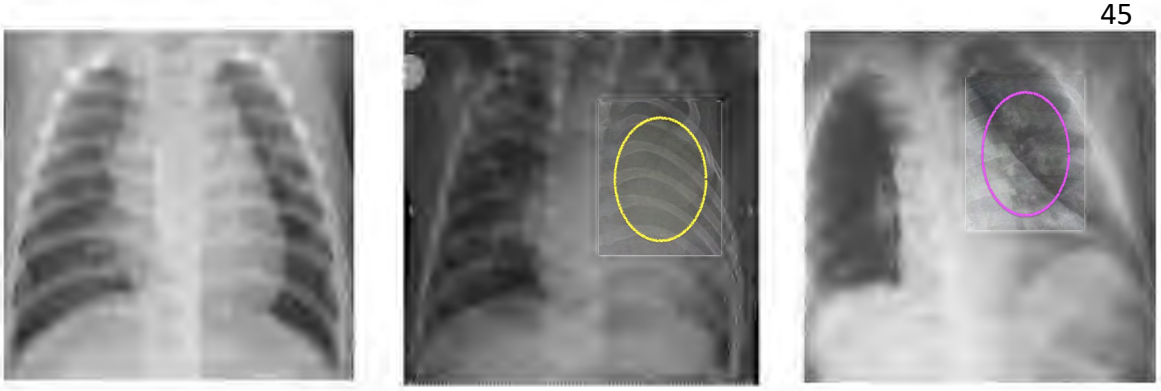
where λ is the weight for l^{MSEL}. Training the proposed model was to optimize loss on the training data. At the beginning of training, the total loss and the learning gradients are dominated by the supervised loss component, that is, the labeled data only. At later stage of training, unlabeled data contributed more than labeled data. These processes are controlled by fine-tuning λ [70]. The detailed steps for learningof the proposed model are shown in algorithm 1. f_θ shared $(\cdot)$ is to learn the low-level features from the medical images. Parameters of the shared ResNet θ_{shared} include weights learned for the operations, namely, pooling operation $f_{pooling}(\cdot)$, convolutional operation $f_{conv}(\cdot)$, and residual operation $f_{Resblock}(\cdot)$.After extracting low-level feature representations from the inputs I used f_θ sup $(\cdot)$ and $f_\theta(\cdot)$ to obtain higher level representations z^{sup} and $z^{unsup}_{Resbloc}$, where z^{sup}_{con} is used to complete COVID-19 classification. In addition, z^{sup} and z^{unsup}_{pool-} were employed to enhance the image representations. Parameters of the supervised ResNet include weights learned for the operations, namely, pooling operation $f^{sup}(.)$, convolutional operation $f^{sup}(\cdot)$, and residual operation $f^{sup}(\cdot)$ while those of the unsupervised ResNet θ_{unsup} consisted of weights learned for the operations, namely,pooling operation $f^{unsup}(\cdot)$, convolutional operation $f^{unsup}(\cdot)$, and residual operation $f^{unsup}(\cdot)$. In the training procedure, the data imbalance was

overcome by assigning more weight w_i to the minority class (COVID-19 class) of samples.Finally, I employed ADAM optimizer to jointly optimize the total loss.

3.1.2 Experiment

3.1.2.1 Dataset

A large-scale of chest x-ray dataset COVIDx [58] was employed to validate the proposed model. It was comprised of 18,543 chest radiography images across 13,725 cases. Examples of chest x-ray images belonging to normal, pneumonia, and COVID-19 classes from COVIDx dataset are shown in Fig. 3.2. When these examples are examined, I can differentiate these images in terms of features shown within areas marked by the blue circle since I can observe some lighter areas indicating COVID-19 infected regions in the blue circle. Additionally, when examining the class distribution between training and testingdata, I noticed that class distribution of the training set was significantly different from that of testing set. Hence the data was rebuilt by splitting the dataset into training and testing datasets that share similar class distributions, where 70% and 30% of data were used for training and testing datasets, respectively. The detailed information of the rebuilt dataset is shown in Table 3.1 for sample distribution.



(a) normal (b) pneumonia (c) COVID-19

Fig. 3.2. Examples of chest radiography images belonging to normal, pneumonia, and COVID-19 classes are shown in (a), (b) and (c), respectively. Yellow circle locates infected regions of pneumonia for subfigure (b) while in subfigure (c) the red rectangle shape of region in the blue circle shows the potential infected areas of COVID-19.

TABLE 3.1. SAMPLE DISTRIBUTION IN DIFFERENT CLASSES FOR TRAINING AND TESTING DATASETS

Dataset	Normal	Pneumonia	COVID-19	Total
Training	6,195	6,708	75	12,978
Testing	2,656	2,876	33	5,565
Total	8,851	9,584	108	18,543

3.1.2.2 *Experimental Settings*

In this experiment, the proposed model performed COVID-19 classification. The key hyper parameters for training the proposed model were: Minibatch size: 256, Number of epochs: 50, Optimizer: Adam optimizer, and Initial Learning rate: 0.1. They were determined by trial and error. Moreover, the details of the model architecture is illustrated in Table 3.2, where the residual block is the standard one [87]. Specifically, the output of the proposed model contains two parts: image class $\phi(z^{sup})$ and a new epresentation z^{unsup}.

COVID-Net[1] [58] is employed as a baseline supervised model to present the state-of-the-art performance of COVID-19 image classification for comparison. Furthermore, the proposed model is compared with SRC-MT [102], that isthe state-of-the-art of semi-supervised learning since it outperformed Π model [70]and mean teacher model [103] in the area of medical image classification.

TABLE 3.2. THE PROPOSED NETWORK ARCHITECTURE.

Name	Description
Input	Medical Images
Shared ResNet	one convolutional layer, 2 residual block, batch normalization, one pooling layer
Supervised ResNet	one convolutional layer, 2 residual block, batch normalization, one pooling layer
Unsupervised ResNet	one convolutional layer, 2 residual block, batch normalization, one pooling layer
Output	image class $\phi(z^{sup})$ and a new representation z^{unsp}

3.1.3 Evaluation Metric

Different evaluation metrics are applied to evaluate the performance of the proposed model. Since the task is a multi-class classification problem, I use accuracy, macro-average Precision (MacroP), macro-average Recall (MacroR), andmacro-average Fscore (MacroF) [44, 104, 105]. Accuracy is calculated by dividing the number of medical images identified correctly over the total number of testing medical images.

$$\text{Accuracy} = \frac{\text{N}_{correct}}{\text{N}_{total}}. \tag{3.13}$$

Macro-average [106] is to calculate the metrics such as Precision, Recall and F-scores independently for each image class and then utilize the average of these metrics. It is to evaluate the whole performance of classifying image classes.

$$\text{MacroF} = \frac{1}{C}\sum_{c=1}^{C} F\text{score}_c \tag{3.14}$$

$$\text{MacroP} = \frac{1}{C}\sum_{c=1}^{C} P\text{recision}_c \tag{3.15}$$

$$\text{MacroR} = \frac{1}{C}\sum_{c=1}^{C} R\text{ecall}_c \tag{3.16}$$

where C denotes the total number of image classes and Fscore_c, Precision_c, Recall_c are Fscore, Precision, Recall values in the c^{th} image class which are defined by

$$\text{Fscore} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.17}$$

where Precision defines the capability of a model to represent only correct image classes and recall computes the aptness to refer all corresponding correct imageclasses:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.18}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3.19}$$

whereas T P (True Positive) counts total number of medical images matched the annotated images. FP (False Positive) measures the number of recognized classes does not match the annotated images. FN (False Negative) counts the number of medical images that does not match the predicted medical images. The ideal case oflearning from imbalanced datasets such as COVIDx is to improve the recall without hurting the precision. However, recall and precision goals are often conflicting, sincewhen increasing the true positive (TP) for the minority class (True), the number of false positives (FP) can also be increased. This reduces the precision [107]. In addition, the confusion matrix is employed to check the detailed performance for each class,

especially on COVID-19 class.

3.2 Calibrated Bagging Deep Learning for Image Semantic Segmentation: A Case Study on Covid-19 Chest X-Ray Image.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes coronavirus disease 2019 (COVID-19) which was first identified in 2019 in Wuhan, compared to RT-PCR tests, medical imaging tests such as chest x-ray (CXR) and computed tomography (CT) are more effective and efficient [111, 112], which is of great help to physicians. For instance, in Italy, the United States, and China, the majority of seriousCOVID-19 cases have been identified through the manifestation characteristics in CTimages [113]. Therefore, effective extraction of COVID-related information on medical images would play an important role to fight against a new round of pandemic caused by COVID mutated variant [114]. Deep learning (DL) played an important role in promoting COVID-related information extraction by COVID-19 infection region segmentation and disease classification through analyzing CXR and CT data [88, 61]. Compared with CT images, CXR images are easier to obtain in radiological inspections. Currently, most of DL models, especially convolutional neural networks (CNN), were employed to classify entire CXR images to detect COVID-19 cases [58, 115]. For example, Hemdan et al. proposed COVIDX-Net to assist radiologists to diagnose COVID-19 based on CXR features [116]. It integrated various deep convolutional neural networks (DCNN) models with different structures, such as DenseNet201 [117], Xception [118], and MobileNetV2 [119]. Sethy et al. integrated different DCNN models with a sup- port vector

machine (SVM) classifier to recognize COVID-19 [120]. In addition,to address the shortcomings of training data, Castiglioni et al. employed transfer deep learning techniques for COVID-19 classification, where the pretrained models were built based on ResNet on ImageNet datasets [121]. Ioannis et al. comprehensively evaluated transfer learning based COVID-19 classification by investigating five DCNN models, including VGG19, MobileNetV2, Inception, Xception, and Inception- ResNetV2 [59]. Similarly, Narin et al. applied three typical pretrained DCNN models (that is, ResNet50, InceptionV3, and InceptionResNetV2) to classify COVID-19 on a small-scale CXR dataset [122]. Moreover, Lucy et al. [72] developed two-path semi-supervised deep learning model to implement COVID-19 classification by using huge amounts of unlabeled data.

Compared with CXR classification, CXR semantic segmentation is a more challenging task, that is, to classify each pixel into predefined classes [123] to recognize region of interests (ROIs) on CXR images, where a few previous work explored this task [124, 125, 126]. Specifically, it seems to not be comprehensively investigated onprediction uncertainty of deep learning models for this task since many DL models focus on performance improvement on this task. However, for safety-critical applications like medical image processing, the prediction uncertainty of DL models isa key evaluation metric on reliability of model predictions, where high prediction uncertainty means low prediction reliability. For example, for COVID-19 applications, applying uncertain predictions to clinical processes would result in disastrous consequences such as missing serious COVID cases or delayed treatments.

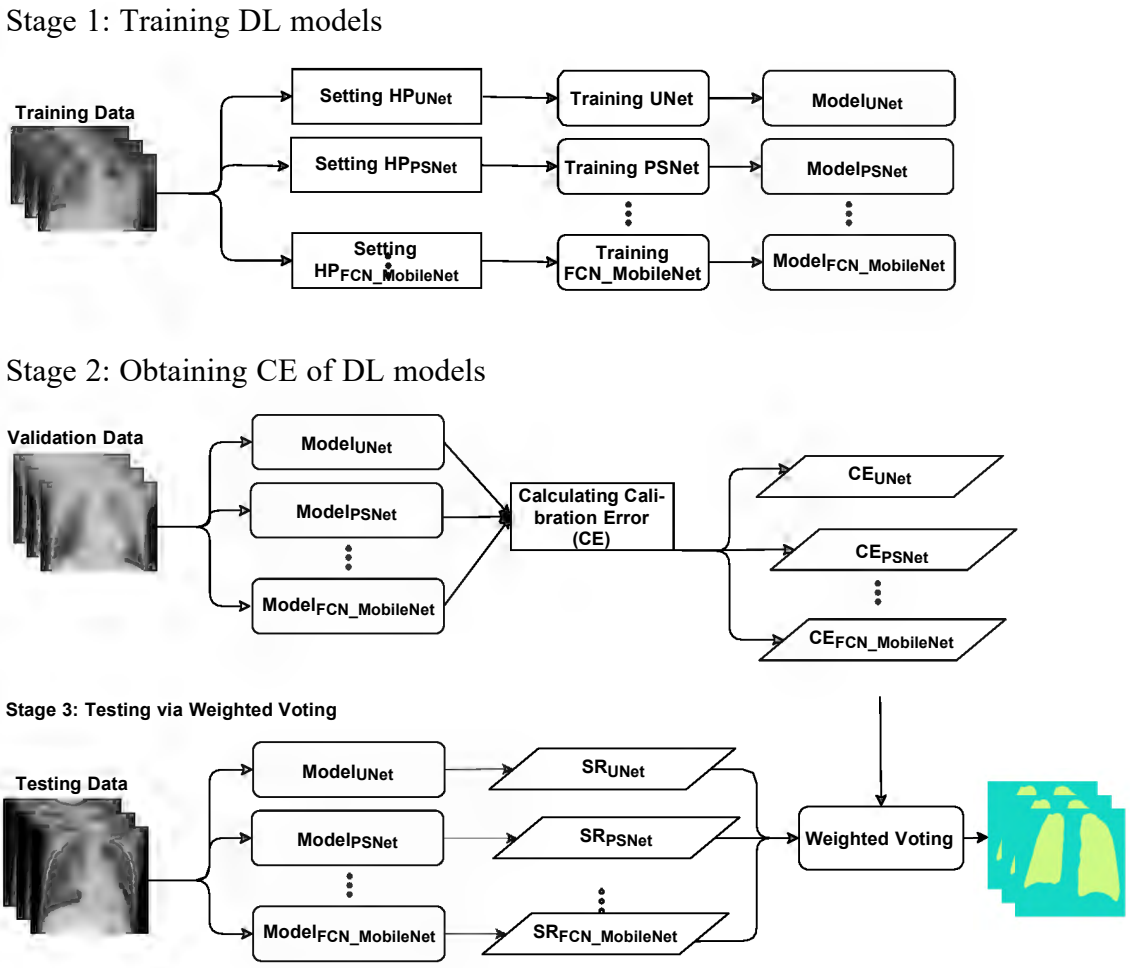


Fig. 3.3. Flow of building and testing calibrated bagging deep learning. SR denotes predictions with individual deep learning model.

This research proposed a novel ensemble deep learning model that integrates bagging deep learning [127] and model calibration [128] to enhance performance of semantic segmentation, as well as reduce prediction uncertainty. It involves three stages:

1) Training multiple state-of-the-art DL models such as fully convolutional networks (FCN) [129], FCN combined with ResNet [87], FCN combined with MobileNet [129], PSPNet [37], and UNet [2] on training CXR datasets.

2) Calculatingcalibration errors to measure prediction uncertainties of these DL models on validation dation CXR datasets, where expected calibration error (ECE) and maximum calibration error (MCE) [128] are employed to estimate the prediction uncertainties.
3) Appealing weighted voting of predictions on testing CXR datasets generated by these DL models to implement calibrated bagging deep learning, where the weight of each DL model is inversely proportional to the calibration error; The proposed modelwas validated on a large-scale CXR dataset to examine its effectiveness. Experimental results demonstrated that the proposed method not only enhanced the performanceof semantic segmentation, but also improved the prediction certainty on CXR data.

The contributions in this study are below.

- Systematically compared performance of various state-of-the-art DL models on semantic segmentation on COVID-19 CXR data with different evaluation metrics. Moreover, the prediction uncertainty of these DL models was investigated by estimating expected calibration error (ECE) and maximum calibration error (MCE).
- Implemented a novel ensemble deep learning model based on model calibration and bagging deep learning, which is to calibrate bagging deep learning models through weighted summation of predictions generated by individual models. The proposed approach was easily implemented and scalable to various tasks.
- Validated the proposed method on semantic segmentation on a large COVID-19 CXR dataset based on different evaluation metrics. Experimental results

demonstrated its effectiveness in improving performance and prediction certainty for semantic segmentation.

3.2.1 Methodology

The method was built based on model calibration [130, 131, 132] andbagging deep learning [127] to enhance COVID-19 image segmentation with higher prediction certainty.

3.2.1.1 Model *Calibration*

The prediction reliability of machine learning models is critical for high risk applications such as medical diagnosis [131, 132] and self-driving [133], which can be formulated as model calibration [130] that refers to the process of adjusting model parameters to make prediction confidence to be accurate estimation of the probability of the correct prediction [128]. Suppose a machine learning model is perfectly calibrated given a class y with the true probability p.

$$P(\hat{y} = y|\hat{p} = p) = p \tag{3.20}$$

where $p \in [0, 1]$ and class labels $y \in \{0, ..., k\}$. It means that when the prediction probability $\hat{p}$ is equal to true probability p, the prediction $\hat{y}$ is the same as the ground true y. Furthermore, the difference between the prediction confidence P and the true probability p is defined as calibration error, that is to estimate model uncertainty. The expected calibration error (ECE) and the maximum calibration error (MCE) were proposed to measure the quality of uncertainty for machine learning models in terms of prediction accuracy [134].

- Expected Calibration Error (ECE). It estimates the calibration error in expectation values with three steps: 1) discretizing the prediction probability regioninto a fixed number of bins; 2) assigning each predicted probability to one of these bins; 3) calculating the difference between the fraction of predictions in

the bin that are correct (accuracy) and the mean of the probabilities in the bin (confidence) by

$$ECE = \sum_{k=1}^{K} \frac{n_k}{N} |acc(k) - conf(k)| \quad (3.21)$$

where n_k is the number of predictions in bin k, N is the total number of samples predicted, and acc(k) and conf(k) denote the accuracy and confidencein the bin k, respectively. It is a weighted average of differences of accuracy vsconfidence in these bins.

- Maximum Calibration Error (MCE). It measures an upper bound of ECE thatis the maximum difference between accuracy and confidence over all predictions across all bins.

$$MCE = \max_{k=1}^{K} |acc(k) - conf(k)| \quad (3.22)$$

In summary, MCE measures the largest calibration gap across all bins, whereas ECE measures a weighted average of all gaps. Both MCE and ECE equal 0 if the model is perfectly calibrated

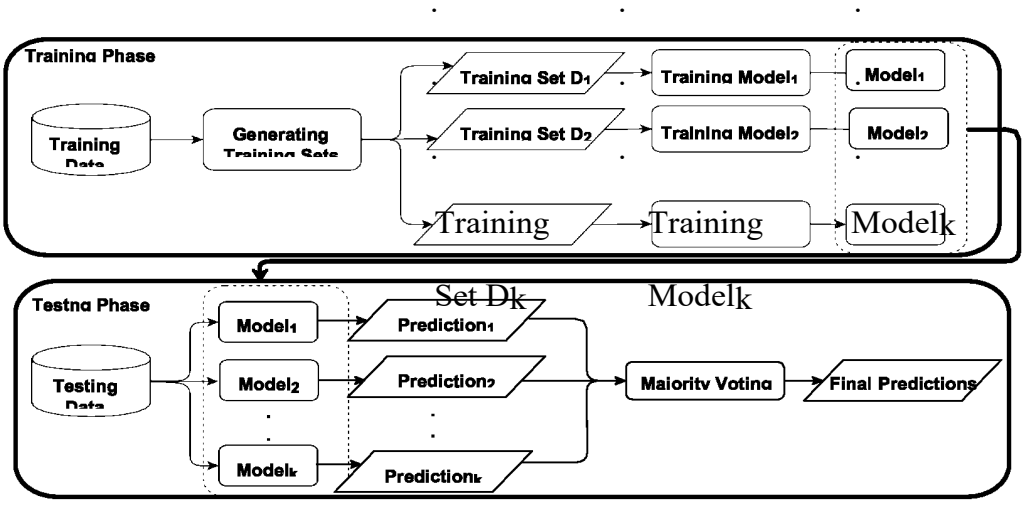


Fig. 3.4. Diagram for building a bagging deep learning model. The model canbe different deep learning models such as convolutional neural networks (CNN) and recurrent neural networks RNN) for different applications.

3.2.1.2 Bagging Learning

Ensemble deep learning combines several individual deep models to improve generalization performance through various ensemble strategies such as bagging and boosting, which integrates the advantages of both deep learning and ensemble learning [127]. Bagging, (or bootstrap aggregating, generates a series of independent sub-sets from training data to build multiple individual predictors to build an ensemble model [135In detail, it generates the bagging samples and passes each bag of samples to base models to build multiple predictors. Then, it is to combine predictions of these multiple predictors with specific strategies such as majority voting. Fig. 3.4 presents a diagram for building and testing bagging deep learning with majority voting, where multiple training sets can be generated by sampling with or without replacement.

3.2.1.3. Proposed Model

A calibrated bagging deep learning model is proposed to enhance generalization performance as well as reduce prediction uncertainty for COVID-19 semantic segmentation. Fig. 3.3 presents the flow for building the proposed approach. It includes three stages: 1) training various state-of-the-art deep learning models suchas UNet [2], PSPNet [37], and MobileNet [129], on an identical training data for COVID-19 image segmentation models, which differs from the standard strategy for bagging learning that is to generate a bag of training sets on original training data; 2)estimating calibration error (CE) for these different models. First, it is to complete COVID-19 semantic segmentation on validation data by running these DL models to obtain prediction probabilities and accuracy. Then, it calculates CE including ECEand MCE to evaluate uncertainties of these DL models; 3) testing via weighted voting

bagging deep learning. I performed calibrated bagging prediction on testing data through implementing weighted voting, where the weights were built with CE ofthese DL models. It assumed that lower CE of DL models meant higher certaintyof these DL models. Moreover, DL models with the higher certainty were assigned with more weights.

3.2.2 Experiment

3.2.2.1 Dataset

COVID-19 chest x-ray dataset was employed to validate the effectiveness of the proposed method. It included 6, 402 images of AP/PA chest x-rays/CT scan with pixel-level polygonal lung segmentation. Each image had a corresponding ground truth with two "Lung" segmentation masks, (rendered as polygons, including the posterior region behind the heart), where the masks included most of the heart, revealinglung opacities behind the heart which may be relevant for assessing the severity of viral infection. Fig. 4.3 shows one example of CXR image and corresponding ground truth. In terms of the example, semantic segmentation on CXR images was to classify pixels in the original image into two classes: Lung, (white region in ground truth, and non-lung, (black region in ground truth). I split the dataset into training (70% data), validation (10% data), and testing (20% data) datasets

3.2.2.2 Experimental Settings

I employed five state-of-the-art individual models as baselines to evaluate performance of semantic segmentation, namely, UNet [2], PSPNet [37], FCN32 [35] (FCN with 32×upsampling), FCN32 ResNet50 (FCN32 combined with ResNet50 [87]), FCN32 MobileNet (FCN32 combined with MobileNet [129]), and an ensemble base-line built based on majority voting, where the ensemble baseline was built based

on bagging learning with these results generated by these five baselines (UNet, PSPNet, FCN32, FCN32 ResNet50, and FCN32 MobileNet). Moreover, key hyper-parameters of these individual models are shown in Table 3.3.

Algorithm 2 Building calibrated bagging deep learning

Require: Training set $D_{training}$ and validation set D_{val}
Ensure: Calibrated bagging deep learning
1: form ← 1 to M do
2: Setting hyper-parameter (HP) for DL_m
3: Training DL_m on $D_{training}$
4: Calculating CE_m of DL_m on D_{val}
5: end for
6: return DL models DL= $\{DL_1, DL_2, ..., DL_M\}$ and corresponding CE = $\{CE_1, CE_2, ..., CE_M\}$

3.2.1.1 Original (b) Groundtruth

Fig. 3.5. (a) Original image (b) Ground truth, an example of CXR image and corresponding ground truth

TABLE 3.3. HYPER-PARAMETERS OF BASELINES FOR COVID-19 IMAGE SEGMENTATION

Model	Learning Rate	Batch Size	Epoch
UNet	1e-3	2	50
PSPNet	1e-3	2	70
FCN32	1e-3	2	50
FCN32 ResNet50 (F32 R50)	1e-3	2	50
FCN32 MobileNet (F32 M)	1e-3	2	50

I implemented two versions of the proposed approach including Ensemble (Weighted Voting (ECE)) and Ensemble (Weighted Voting (MCE)). Ensemble (WeightedVoting (ECE)) is a weighted bagging learning method, where the weights are obtained by calculating expected calibration error (ECE). Similarly, Ensemble (Weighted Voting (MCE)) is a weighted bagging learning method, where the weights are obtained by calculating maximum calibration error (MCE). Moreover, I combined the predictions of Ensemble (Majority Voting (MV)), Ensemble (Weighted Voting (ECE)), and Ensemble (Weighted Voting (MCE)) by majority voting to build Ensemble (MajorityVoting + ECE + MCE (MVEM)).

3.2.3 Evaluation Metric

Various evaluation metrics were employed to evaluate the performance of the proposed model, which included accuracy, F1score, sensitivity, and specificity. Accuracy was calculated by dividing the number of pixels identified correctly over the total number of pixels in chest x-ray images.

$$\text{Accuracy} = \frac{\text{N}_{correct}}{\text{N}_{total}}. \tag{3.23}$$

$$\text{Fscore} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3.24}$$

where Precision defines the capability of a model to represent only correct pixelsand Recall computes the aptness to refer all corresponding correct pixels.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3.25}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{3.26}$$

.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.27}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3.28}$$

whereas T P (True Positive) counts the total number of pixels that matches the annotated pixels of RIOs. FP (False Positive) measures the number of pixels that do not belong to RIOs, but are recognized as pixels of RIOs. FN (False Negative) counts the number of pixels of RIOs are recognized as those that do not belong to RIOs. The main goal for binary classification is to improve the recall without hurting the precision. However, recall and precision goals are often conflicting, since when increasing the true positive (TP) for the minority class (True), the number of false positives (FP) can also be increased. This will reduce the precision [107].

Moreover, I employed sensitivity and specificity to evaluate performance of semantic segmentation [136], where the sensitivity measures how good a test is at detecting the RIOs while the specificity refers to how good a test is at avoiding false alarms. Finally, I employed expected calibration error $(ECE)^2$ and MCE to calculate the calibration errors [128] for evaluating the prediction uncertainty, where ECE and MCE were defined as equations (1) and (2), respectively. The lower the ECE and MCE, the higher the prediction certainty.

3.3 Proformer-Based Ensemble Learning for Gene Expression Prediction

DREAM Challenges are widely acknowledged as a pioneering force in the field of biomedical research [145], with a primary objective of encouraging transparent, cooperative, and open scientific inquiry in biomedicine. These challenges strive to enhance techniques in systems biology by tackling intricate and groundbreaking problems in biology and biomedicine. A staggering number of individuals, exceeding 30,000 from different parts of the world, have benefited from these challenges, encompassing individuals with diverse backgrounds and multidisciplinary interests. The utilization of open-source code and data plays a significant role in driving the success of these challenges. Moreover, it is equally essential to highlight the biological insight and translatability of the algorithms used to address the queries presented in the challenges.

Comprehending the cis-regulatory logic of the human genome is crucial as it would offer insights into the origins of various diseases. Nonetheless, learning models from human data faces many challenges, such as limitations in the diversityof sequences, extensive repetitive DNA, a vast number of cell types that interpret regulatory DNA differently, limited reporter assay data, and substantial technical biases. The DREAM Challenges 2022 [146] have addressed these issues by creating high-throughput measurements of cis-regulatory activity for millions of randomly generated promoters in the single-cell organism Yeast. The level of expression produced by each promoter sequence is assessed using a fluorescent reporter gene that is regulated by the promoter. The number of randomly generated promoter sequencesis so vast that it is comparable in complexity to the entire human genome, providing

unprecedented opportunities to learn the numerous parameters for comprehending gene regulation. As the cis-regulatory logic of both human and Yeast systems is based on similar principles, it is hoped that the model architectures learned from the yeast data would be promising for developing models for the human genome. During this competition, participants will receive expression measurements of millions of promoter sequences that are randomly generated to train machine learning modelsthat predict gene expression from the sequences.

From a machine learning perspective, predicting gene expression can be seen as a regression problem, where the gene sequences serve as inputs and the corresponding expression levels are the outputs [147]. In recent years, deep learning techniques have been utilized to improve gene expression prediction, with deep convolutional neural networks (CNNs) currently achieving the best performance for predicting gene expression from DNA sequences in both human and mouse genomes [148, 149, 150, 151]. For example, ExPecto [148], a deep learning-based framework, can accurately predict tissue-specific transcriptional effects of mutations, including rare or unobserved ones, from a DNA sequence. Kelley et al. trained deep CNNs simultaneously on multiple genomes and applied it to large compendia of human and mouse data, improving the accuracy of gene expression prediction on variant sequences [149]. Agarwal et al. applied deep CNNs to predict gene expression levels based solely on genome sequence, and achieved surprising success in explaining up to 59% and 71% of variationin steady-state mRNA levels in human and mouse, respectively, using only promotersequences and mRNA stability-related features [150]. Despite the promising resultsof deep learning-based gene expression prediction, there have been few attempts

to leverage the power of deep learning with millions of gene sequences.

In this chapter, I proposed an ensemble deep learning models based on Pro-former [5] to successfully perform gene expression prediction through training and testing on millions of gene sequences. Proformer is an end-to-end Transformer [43] encoder as seen in Fig. 3.7, to predict the expression values from DNA sequences.

It utilizes a Macaron-like Transformer encoder architecture with two half-step feed forward (FFN) layers placed at the beginning and end of each encoder block, along with a separable 1D convolution layer inserted after the first FFN layer and before the multi-head attention layer. Proformer achieved 3rd place in the DREAM Challenges 2022 for gene expression prediction. To further improve its performance, I introduced ensemble strategies by building multiple Proformers and combining their predictions using various weighted strategies. Experimental results demonstrated that the proposed method not only enhanced the performance of Proformer effectively regarding four evaluation metrics, but also achieved higher performance than the winning of the DREAM Challenges 2022 in terms of values of Score Spearman and Spearman.

The contributions in this study are below:

- The proposed novel approach involves an ensemble model consisting of various end-to-end transformer encoders with different architectures. The model predicted the gene expression values of promoter sequences by generating predictions using individual models and combining them through averaging weighted summation.
- I evaluated the proposed model on thousands of randomly generated promoter sequences in Yeast using various evaluation metrics. The experimental results

demonstrated the effectiveness of the approach in improving the performance of existing state-of-the-art methods through effectively learning feature representations of promoter sequences.

3.3.1 Methodology

The proposed method was built based on Proformer [5] and ensemble deep learning [127] to enhance gene expression prediction.

3.3.1.1 Proformer

Proformer is a deep learning model that uses the Transformer encoder architecture to predict expression values from DNA sequences in an end-to-end manner [5]. Its unique design includes a Macaron-like Transformer encoder architecture with two half-step feed forward (FFN) layers at the beginning and end of each encoder block, respectively. In addition, a separable 1D convolution layer is inserted after the first FFN layer and before the multi-head attention layer. Proformer utilizes sliding k-mers from one-hot encoded sequences that are mapped onto a continuous embedding, along with learned positional and strand embeddings (forward strand vs reverse complement strand) as the sequence input. The model also uses multiple expression heads, which predicts expression values for each head and then averages the predictions of all heads to obtain the final predicted expression value. It achieves significantly better than conventional methods, such as using a global pooling layer as the output layer for regression tasks. In summary, Proformer provides a new and effective approach to learn and characterize how cis-regulatory sequences determine expression values.

3.3.1.2 Ensemble Deep Learning

Ensemble deep learning combines several individual deep models to improve generalization performance through various ensemble strategies such as bagging and boosting, which integrates the advantages of both deep learning and ensemble learning [127]. Bagging (or bootstrap aggregating) generates a series of independent sub-sets from training data to build multiple individual predictors to build an ensemble model [135]. In detail, it generates the bagging samples and passes each bag of samples to base models to build multiple predictors. Then, it is to combine predictions of these multiple predictors with specific strategies such as majority voting.

3.3.1.3 Proposed Model

I proposed a Proformer-based ensemble model to enhance gene expression prediction, where the flow of building the proposed method is shown in Fig. 3.6. The process began with the preprocessing of data, which involved several steps such as gene reverse complement, gene masking, and K-mer representation. The gene reverse complement was to reverse the gene order of original gene sequences. The gene masking was to mask individual gene randomly. The K-mer representation was to map one-hot encoded sequences onto a continuous embedding, combined with the learnt positional embedding and strand embedding, (forward strand vs reverse complement strand, as the sequence input. The preprocessed data was then used to train multiple Proformers, each with a unique configuration, where a Proformer consisted of four Macaron Encoders [5]. Then these training Proformers were employed for gene expression prediction on the test gene sequences to generate gene expression level (GEL). The final GEL was produced by merging the different GELs obtained from the

various Proformers using a weighted summation, where the weight can be built bad on evaluation metrics.

3.3.2 Experiment

3.3.2.1 Dataset

Dream Challenges 2022 provided millions of randomly generated promoters in the single-cell yeast organism [146]. In detail, it consisted of more than 6 million promoter sequences, each having identical leading 17 nucleotides, identical trailing 13 nucleotides, and a corresponding expression level ranging [0.0, 17.0] with the distribution shown in Fig. 3.8. A fluorescent reporter gene regulated by a promoteris was used to measure the expression level produced by each promoter sequence. The number of randomly generated promoter sequences was so vast that it was comparable in complexity to the entire human genome. This provided an unparalleled opportunity togain insights into the numerous parameters necessary for comprehending gene regulation. Since the cis-regulatory logic of both humans and yeast follows comparable principles, the model architectures learned from yeast data can offer guidance on howto construct models for the human genome.

Furthermore, the data was preprocessed by trimming the identical nucleotides in each promoter sequence and by padding the sequences that were less than 100 nucleotides on the left and right sides with a specific letter K, which was to ensure a standard promoter length of 100 nucleotides for training and testing. Moreover, the reverse complemented sequences, which was formed by interchanging A and T and interchanging C and G in the promoter sequence, were concatenated with the original sequences, (after trimming and padding, as the input for model training. Thus, the

total length of the input sequences was 200 nucleotides. The expression levels were standardized to the mean of zero and a standard deviation of one for better model generalization performance and faster convergence.

3.3.2.2. Experimental metrics

Various evaluation metrics were employed to evaluate the performance of the proposed model, namely, PearsonR, Spearman, score PearsonR, score Spearman.

- PearsonR: PearsonR's correlation measures the strength of the linear correlation of the standardized slope of a simple linear regression line (fit) [152].

$$\rho_p = \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(xi - x)^2 (yi - y)^2}}$$

ρ_p = Pearson correlation coefficient

x_i = Values of the x-variable in a sample

$\bar{x}$ = Mean of the values of the x-variable

y_i = Values of the y-variable in a sample

$\bar{y}$ = Mean of the values of the y-variable

where the x-variable is the gene in the gene sequence while the y-variable isthe gene expression.

- Spearman: Spearman's correlation ρ_s determines the strength and direction of the monotonic relationship between two ranked variables [152].

$$\rho_s = 1 - \frac{6\,(D_i)^2}{n(n^2 - 1)}$$

Where $D_i^2 = R(x_i) - R(y_i)$

D_i^2 = Squared of the difference between the two ranks of each x and y observation

R = rank for each observation in a sample

x_i = Values of the x-variable in a sample

y_i = Values of the y-variable in a sample

n = Number of observations

- Score PearsonR: A weighted measure of the strength of the linear correlation of the standardized slope of a simple linear regression line (fit). In this study, the mean square error is used as weight [153].

$$\rho_{sp} = \frac{\sum_{i=1}^{n}[w_i(x_i - x)(y_i - y)]}{\sum_{i=1}^{n}(w_i(x_i - .x)_2)\sum_{i=1}^{n}(w_i(y_i - .y)_2)}$$

ρ_{sp} = Score PearsonR correlation coefficient

n = Number of observations

- Score Spearman: a weighted rank measure of correlation that weights the distance between two ranks using a linear function of those ranks, giving more importance to higher ranks than lower ones. In this study, the mean square error was used as weight [153].

$$\rho_s = \frac{1 - 6\sum_{i=1}^{n} W_i(D_i)^2}{n(n^2 - 1)}$$

ρ_{ss} = Score Spearman correlation coefficient

w_i = Mean square error

n = Number of observations

3.3.3 Experimental setup

I proposed a novel ensemble of end-to-end Transformer encoder architectures that leveraged the self-attention mechanisms and could handle long-range correlations between the input-sequence items to predict the expression values from millions of DNA sequence. Specifically, the values of evaluation metrics including Spearman, PearsonR,

Score Spearman, and Score PearsonR were employed as weight to implement the weighted sum during the prediction and average it to obtain the final predicted gene expression values. The key hyper-parameters of these individual models are shown in Table 3.4.

TABLE 3.4. HYPER-PARAMETERS OF BASELINE MODELS

Parameters	Model 1	Model 2	Model 3	Model 4
Attention Heads	4	8	8	8
Encoder Blocks	4	4	4	4
Expression Heads	NA	1	32	32
Batch size	512	512	512	512
Masking	NA	NA	5%	5%
Learning Rate	1e-3	1e-3	1e-3	1e-3
Epochs	20	20	20	20

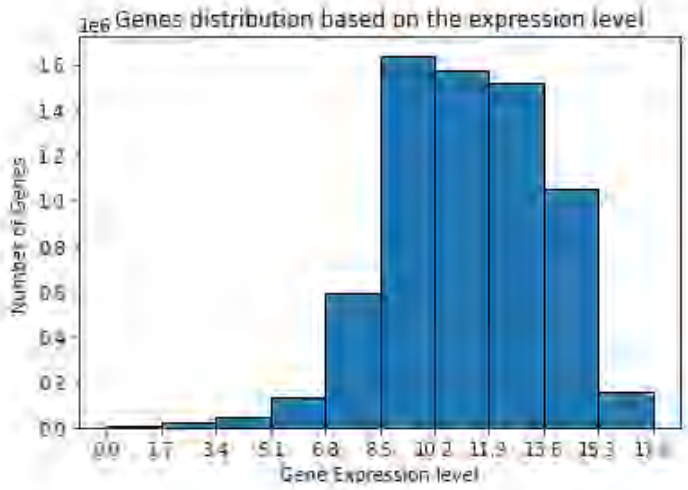


Fig. 3.6. Distribution of genes expression levels.

The training losses consisted of the mean squared error between the expression values (y mse), and the reconstruction loss (y recon) which helps stabilize the training process especially with larges models, where five percent of the nucleotides were randomly masked and predicted. The mean of the prediction of all heads for all predicted expression value from for each head was used as the final predicted expression value.

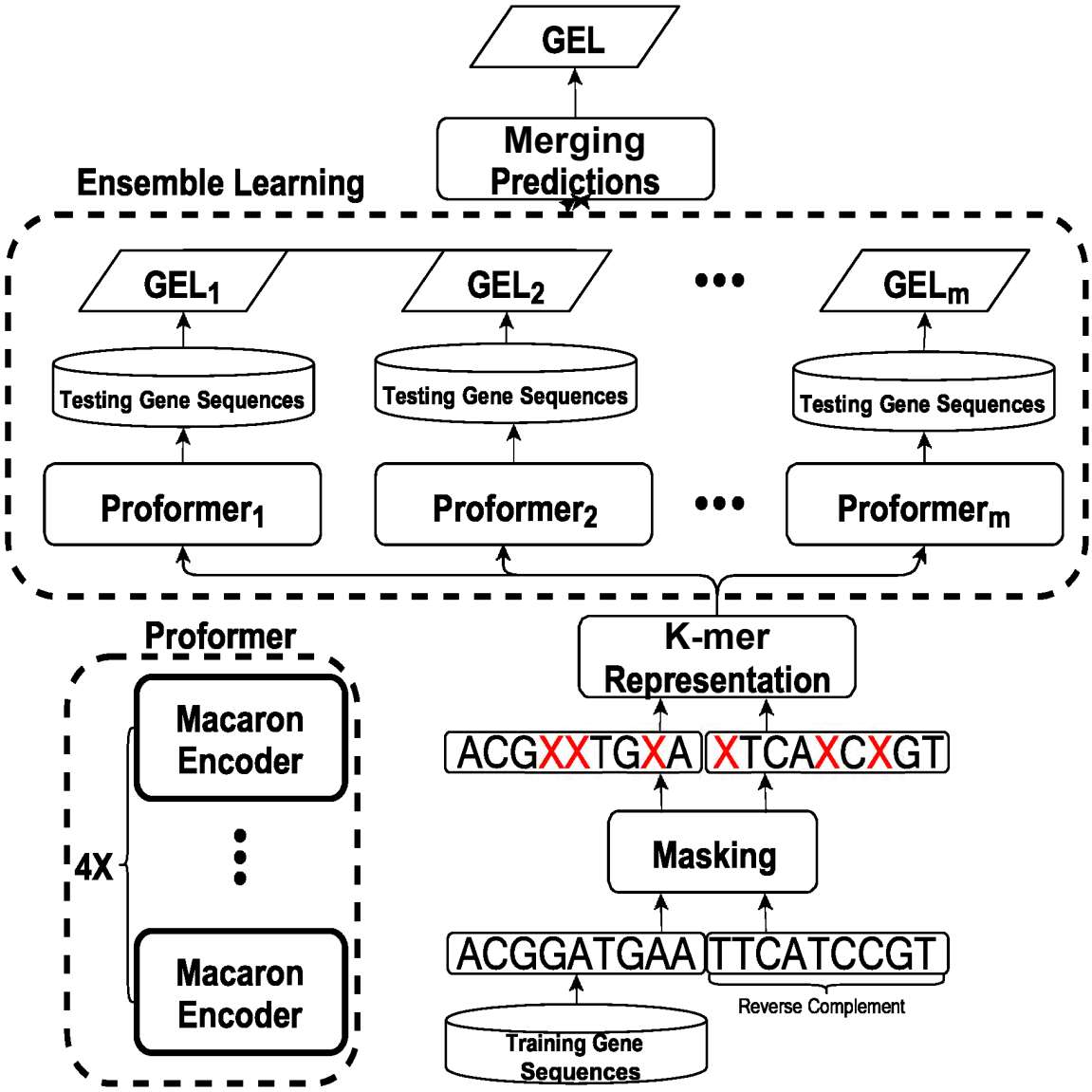


Fig. 3.7. Flow of the proposed ensemble Proformer models. It starts with data preprocessing including gene reverse complement, gene masking, and K-mer representation. Outputs of the data preprocessing are employed to train multiple Proformer with different setups, where the Proformer is composed of four Macaron Encoder [5]. These Proformer perform gene expression prediction on the testing gene sequences to conduct corresponding gene expression level (GEL). Finally, these different GELs are merged through averaging weighted summation of predictions to produce the final GEL.

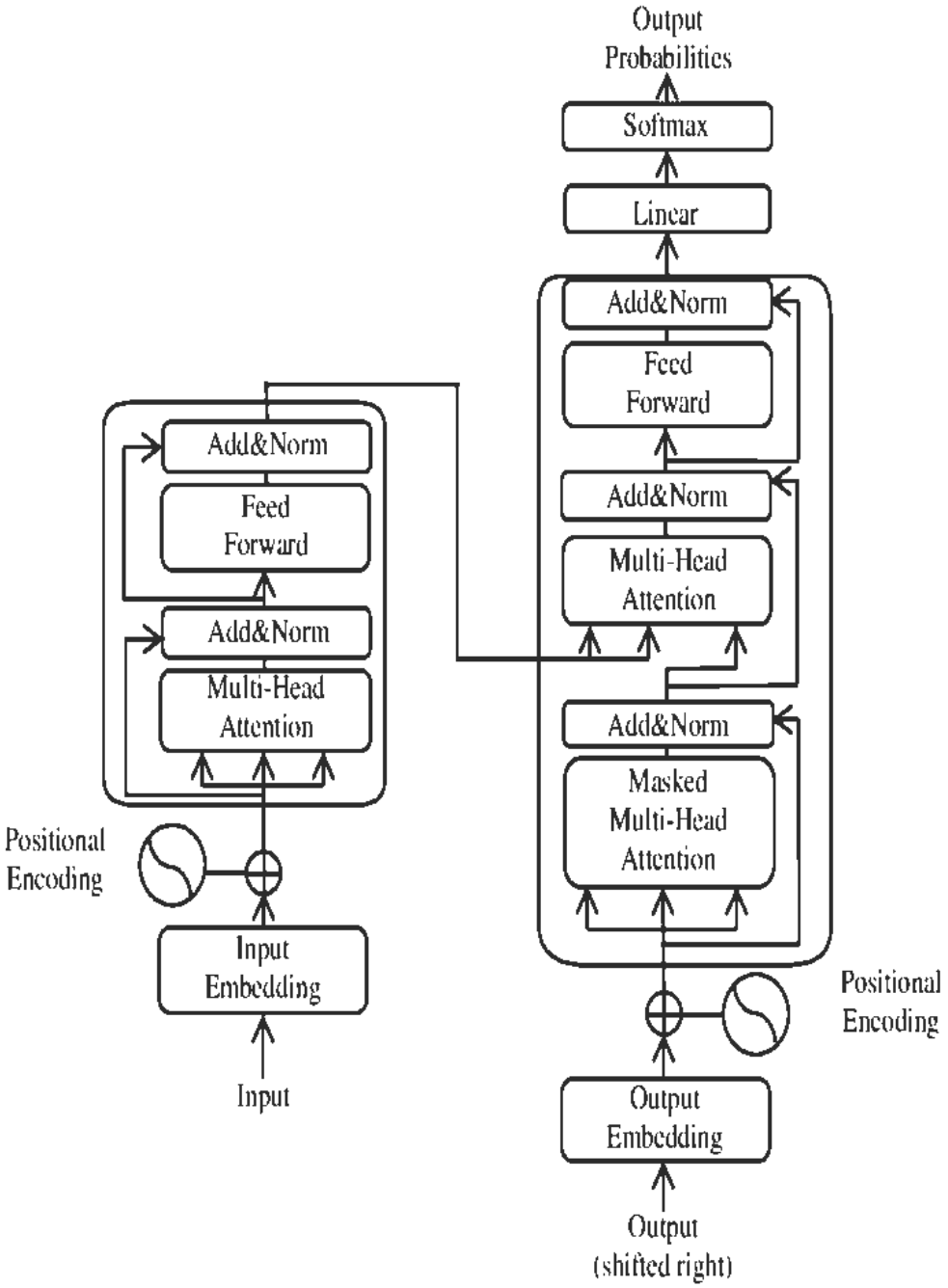


Fig. 3.8. The transformer model architecture[43]

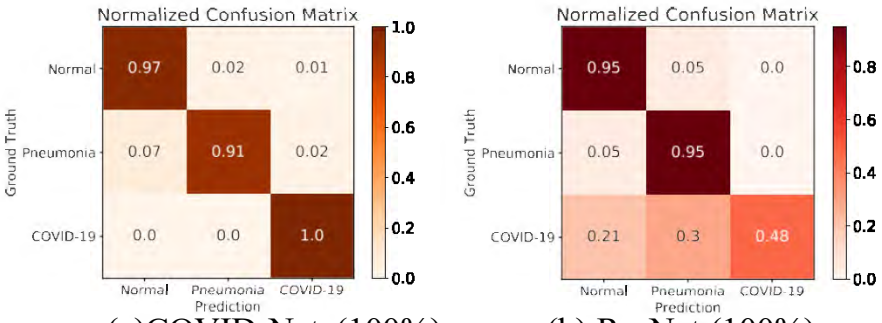
CHAPTER 4

4. FINDINGS

4.1 Experimental Findings for Semi-Supervised Learning For Covid-19 Image Classification Via Resnet

4.1.1 Experimental Results

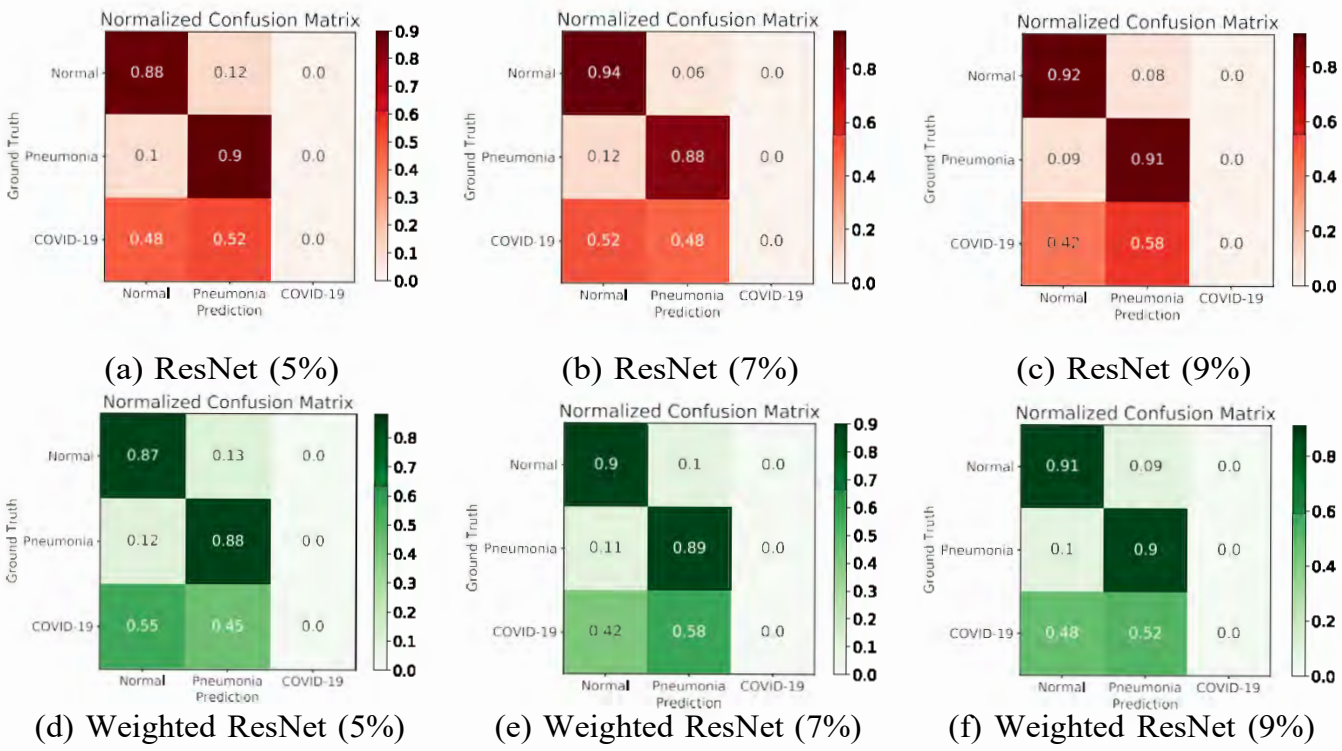
The proposed model performance in four steps was evaluated. The first step was to examine the performance of supervised learning baselines, which was to prove if ResNet was a reasonable supervised model for COVID-19 image classification. A competitive supervised baseline is useful to compare the proposed semi-supervised model in order to present the effectiveness of the proposed model. Furthermore, I checked whether fewer labeled data would lead to lower performance. The second step was to compare the proposed model with state-of-the-art semi-supervised learning.



(a)COVID-Net (100%) (b) ResNet (100%)

Fig. 4.1. Comparison of confusion matrix generated by COVID-Net (100%) and ResNet (100%).

The third step was to examine whether the hyper-parameter setting would affect the performance of the proposed model significantly. Finally, the inability of the proposed model to classify certain COVID-19 cases is discussed.



(a) ResNet (5%) (b) ResNet (7%) (c) ResNet (9%)

(d) Weighted ResNet (5%) (e) Weighted ResNet (7%) (f) Weighted ResNet (9%)

Fig. 4.2. Comparison of confusion matrix generated by different ResNets training on different ratios of labeled data. Weighted ResNet is built by assigning moreweight to COVID-19 class during training for overcome the challenge of data imbalance.

4.1.1.1 Supervised Learning for COVID-19 Classification

Table 4.1 presents the comparison of supervised baselines built with ResNet. I observed that ResNet (100%) outperformed COVID-Net (100%) when comparing accuracy, macro-average precision, and macro-average Fscore. It means that ResNetis a competitive supervised baseline for COVID-19 image classification. To learn from fewer labeled data, I only focused on the cases of five percent, seven percent, and nine percent labeled data since the labeled data would be very scarce in the medical domain [4] during the early stage of a global pandemic such as COVID-19 outbreak.

TABLE 4.1. COMPARISON ON SUPERVISED BASELINE PERFORMANCE. RESNET IS TRAINED ON DIFFERENT RATIOS (%) OF LABELED X-RAY IMAGES. WEIGHTED RESNET IS BUILT BY ASSIGNINGMORE WEIGHT TO COVID-19 CLASS DURING TRAINING FOR OVERCOMING THE CHALLENGE OF DATA IMBALANCE

DL	Accuracy (%)	MacroP (%)	MacroR (%)	MacroF (%)
COVID-Net (100%)	93.98	72.70	96.05	77.33
ResNet (100%)	94.68	90.52	78.87	83.04
ResNet (5%)	87.59	58.23	58.50	58.30
ResNet (7%)	89.27	59.51	59.91	59.68
ResNet (9%)	90.10	60.05	60.44	60.24
Weighted ResNet (5%)	86.45	57.63	57.95	57.78
Weighted ResNet (7%)	88.70	59.15	59.54	59.31
Weighted ResNet (9%)	89.36	66.16	60.46	60.78

4.1.2 Supervised Learning for COVID-19 Classification

It was observed that the classification accuracy can be improved by increasing the labeled data to train ResNet. Meanwhile the performance such as accuracy and MacroF was reduced significantly when comparing with ResNet (100%), which demonstrated that more labeled data was imperative for building high performance supervised models. Moreover, it was observed that weighted ResNet could not improve the performance since inappropriate weight might be assigned to different classes.

Fig. 4.1. indicates that ResNet (100%) could be a promisingsupervised baseline model when compared to COVID-Net in terms of the accuracies on the normal and pneumonia classes. For the COVID-19 class, ResNet was lower thanCOVID-Net since COVID-Net employed transfer learning to enhance performance. To check the performance for each class when learning on fewer labeled data, the detailed performance with confusion matrix shown in Fig. 4.2 is presented. With lowratios of

labeled training data to train models, ResNet cannot recognize COVID-19images effectively, which is due to insufficient COVID-19 labeled samples. In thetraining sets of these cases, only a few of images are for the COVID-19 class. Forexample, in the case of ResNet (five percent), only three images for COVID-19 class were present in the training data, which means that most of training images were for the classes of Normal and Pneumonia. Learning on this data will lead to classification bias. Weighted ResNet was not sufficient to enhance the performance, which means that even more weight assigned to COVID-19 class was not enough to overcome the lack of labeled samples to learn distinct features to differentiate COVID-19 patients from non-COVID-19 patients on x-ray images with supervised learning

4.1.3 Comparing the Proposed Model with State-of-the-art Semi-Supervised Learning

In this section, I examine whether the proposed model was able to effectively identify COVID-19 samples by training on very limited number of annotated images. Table 4.2 presents the comparison of classification performance between SRC-MT and the proposed model (SSResNet). Overall accuracies of SRC-MT are better than those of the proposed model. However, when only five percent labeled samples were usedfor training, MacroF of the proposed model was higher than that of SRC-MT, which indicates that the proposed model was more effective in detecting COVID-19 samples and can detect COVID-19 samples with higher performance. It means that compared to SRC-MT, the unsupervised path could enhance the data representation for improving COVID-19 classification more effectively.

In addition, detailed performance of each class with confusion metrics shown in

Fig. 4.3 are examined. I observed that the accuracy of recognizing COVID-19 by the proposed model is higher than that of SRC-MT, which means SSResNets can learn more effective features from unlabeled data to recognize COVID-19 samples. Furthermore, with the increased ratios of labeled data, the accuracies of recognizing COVID-19 is enhanced significantly. It means that the unsupervised path can enhance the representations of images to improve the classification. In other words, unlabeled data contributed to increasing the COVID-19 classification performance significantly by enhancing the image representations with the unsupervised path ofthe SSResNet.

4.1.3.1 Hyper-Parameter Setting

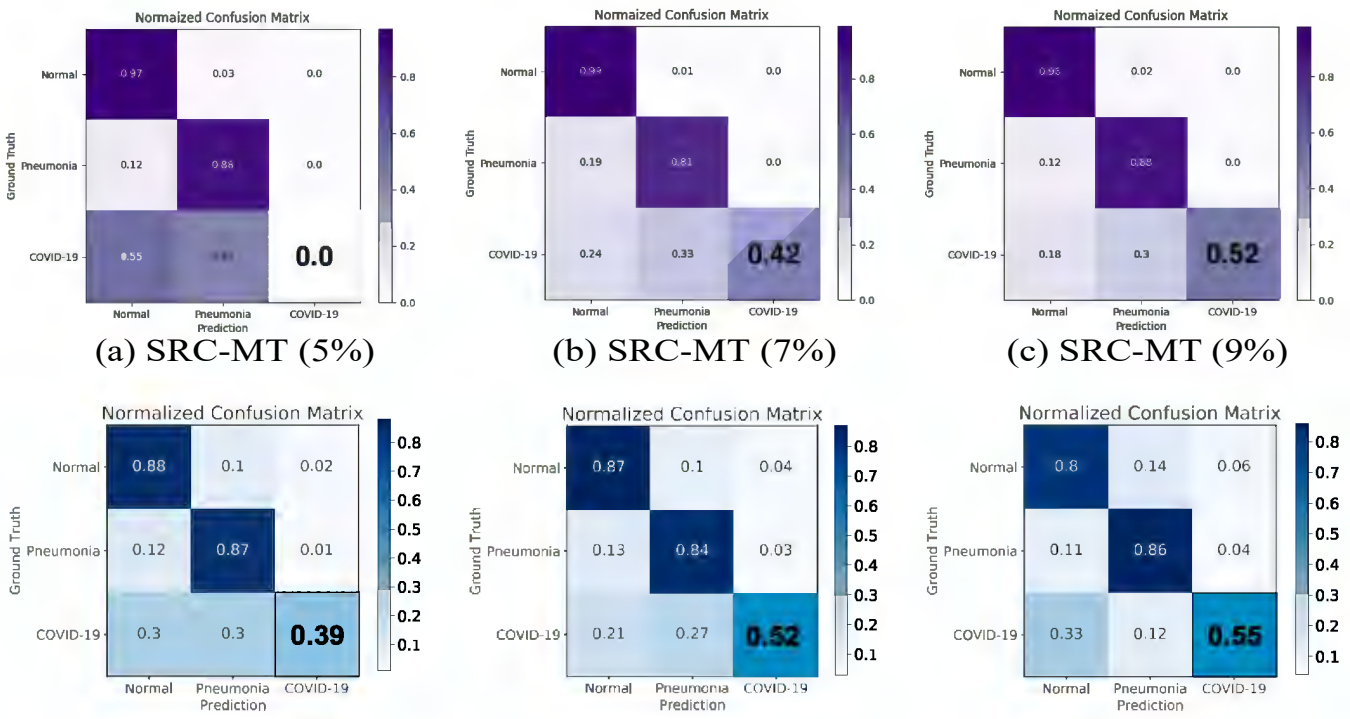
In addition to examining the performance comparison between the proposed models and baselines, the sensitivity of the proposed model to hyper-parameters was determined. There are various hyper-parameters involved in the learning procedure of the proposed model. Here, the class weight was checked since different weights would lead to different performance of recognizing COVID-19 samples. Table 4.3 shows the comparison results for different weights of three classes. It is observed that different weight resulted in significant differences of the performance when examining the values of accuracy. On the other hand, compared to accuracy and MacroP, MacroRand MacroF were less sensitive to the weight of COVID-19 class. Generally, the weight for COVID-19 class was delicately selected to obtain the optimal performance.

TABLE 4.2. COMPARING PERFORMANCE BETWEEN SRC-MT AND OUR MODEL (SEMI-SUPERVISED RESNET (SSRESNET))

Semi-supervised Model	Accuracy (%)	MacroP (%)	MacroR (%)	MacroF (%)
SRC-MT (5%)	90.67	61.08	60.75	60.59
SRC-MT (7%)	89.82	89.92	74.13	78.95
SRC-MT (9%)	92.79	93.61	79.15	84.15
Our model	Accuracy (%)	MacroP (%)	MacroR (%)	MacroF (%)
SSResNet (5%)	84.95	61.18	66.76	62.41
SSResNet (7%)	84.21	63.67	67.85	62.83
SSResNet (9%)	81.79	59.34	70.99	59.19

4.1.3.1.1 *Error Analysis*

Fig. 4.4 presents three COVID-19 samples that are classified into Normal, COVID-19, and Pneumonia classes, respectively. x-ray images of COVID-19 patients shows various features for different stages of COVID-19 patients[2]. At the early stage of COVID-19 patients, x-ray images cannot present significant features (Fig. 4.4 (a)) that can be used to differentiate COVID-19 and Non COVID-19 patients, which leadsto the incorrect classification result for the sample. It is consistent with the expectation that x-ray images are not ideal evidence to support diagnosis of COVID-19 for the patients at the early stage. However, with development of COVID-19, x-ray images are able to present obvious features such as multifocal lung airspace opacities, nodules and consolidation(Fig. 4.4 (b)), which contributes to the correct classification result. Unfortunately,if the patients are at the late stage of COVID-19, x-ray images present lobar diffused consolidation (See Fig. 4.4 (c)) that is similar to features of pneumonia. These features were confusing to the proposed model and lead to the incorrect result



(a) SRC-MT (5%) (b) SRC-MT (7%) (c) SRC-MT (9%)

(d) SSResNet (5%) (e) SSResNet (7%) (f) SSResNet (9%)

Fig. 4.3. Comparison of confusion matrix generated with SRC-MT and SSResNets trained on different ratios of labeled data.

(a)Normal (b) COVID-19 (c) Pneumonia

Fig. 4.4. COVID-19 samples classified into Normal, COVID-19, and Pneumonia classes, are shown in (a), (b), and (c) respectively. The blue circles locate the infected regions of COVID-19.

TABLE 3.5. COMPARING PERFORMANCE WITH DIFFERENT CLASS WEIGHTS. C_1, C_2, AND C_3 ARETHE WEIGHTS OF NORMAL CLASS, PNEUMONIA CLASS, AND COVID-19 CLASS, RESPECTIVELY.

	5% Labeled Data			
Clas Weights	Accuracy	MacroP	MacroR	MacroF
$c_1 = 1, c_2 = 1, c_3 = 2$	84.95	61.18	66.76	62.41
$c_1 = 1, c_2 = 1, c_3 = 5$	78.82	58.18	67.38	56.79
$c_1 = 1, c_2 = 1, c_3 = 10$	66.78	57.24	66.70	50.77
	7% Labeled Data			
Class Weight	Accuracy	MacroP	MacroR	MacroF
$c_1 = 1, c_2 = 1, c_3 = 2$	85.90	65.75	64.62	64.84
$c_1 = 1, c_2 = 1, c_3 = 5$	84.21	63.67	67.85	62.83
$c_1 = 1, c_2 = 1, c_3 = 10$	79.31	58.57	67.71	59.39
	9% Labeled Data			
Class Weight	Accuracy	MacroP	MacroR	MacroF
$c_1 = 1, c_2 = 1, c_3 = 2$	87.28	70.32	62.93	64.65
$c_1 = 1, c_2 = 1, c_3 = 5$	84.69	60.91	66.79	61.86
$c_1 = 1, c_2 = 1, c_3 = 10$	81.79	59.34	70.99	59.19

for the sample shown in Fig. 4.4 (c). In summary, in terms of samples shown in Fig. 4.4, the proposed model was effective for the patients who were in the development of COVID-19 rather than those at the early stage or late stage of such disease.

4.1.4 Discussion

Deep learning technique has shown its power on classification of COVID-19. Ghoshal et al. [88] proposed a Bayesian convolutional neural network to estimate the diagnosis uncertainty in COVID-19 prediction, where the dataset includes 70 lung X-ray images of patients with COVID-19 from an online COVID-19 dataset [89], and non-COVID-19 images from Kaggle's Chest x-ray data (Pneumonia). Narinet al. [90] is to detect COVID-19 infection from X-ray images through comparing three different deep learning models, namely, ResNet50, InceptionV3, and Inception-ResNetV2. The evaluation results showed that the ResNet50 model outperformed other

two models. Zhang et al. [91] also utilized ResNet to complete COVID-19 classification on X-ray images and estimated an anomaly score to optimize the COVID-19 score for the classification. In addition, Wang et al. [58] propose COVID-Net to detect COVID-19 cases using X-ray images. In general, most current studies use X-ray images to differentiate between COVID-19 and other pneumonia and healthy subjects. However, with limited number of COVID-19 images, it is insufficient to evaluate the robustness of the methods and also poses questions to the generalizability. Semi-supervised deep learning has attracted lots of attention since it has the strong ability to generalize the model performance through learning on labeled data and unlabeled data [67, 68, 69, 70]. Generally, it is to train the deep neural networks by jointly optimizing the standard supervised classification loss on labeled samples and an unsupervised loss on unlabeled data [67, 70]. The rationale of these semi-supervised learning models is to enrich the supervision signals by exploiting the knowledge learned on unlabeled data [92] or regularize the network by enforcing smooth and consistent classification boundaries [69]. Regarding COVID-19 research such as COVID-19 image classification and image segmentation, semi-supervised learning is employed to resolve lack of labeled data [93, 94, 95, 96, 97, 98]. However, for COVID-19 image classification, these studies [93, 94, 95] have not comprehensively examined the model performance on a large-scale of X-ray image dataset such as COVIDx [58] by comparing with the state-of-the-art, especially for the case of very few labeled data such as less than ten percent labeled data. This research proposed a semi-supervised deep learning model for COVID-19 image classification and checked the model performance systematically on the COVIDx [58] dataset.

- Computed Tomography Scan (CT-SCAN): a CT scan is a medical imaging technique that combines X-ray technology with computer processing to generate detailed cross-sectional images of the body [47]. It uses a series of X-ray beams taken from different angles to create a 3D representation of internal structures. CT scans are particularly useful for visualizing bones, organs, soft tissues, and blood vessels. They can provide information about the size, shape,density, and location of various structures, aiding in the diagnosis and monitoring of conditions such as fractures, tumors, infections, and internal bleeding.
- Magnetic Resonance Imaging (MRI): MRI is a non-invasive imaging technique that uses a powerful magnetic field and radio waves to generate detailed images of the body's internal structures [47]. It provides highly detailed images of organs, tissues, and other structures without using ionizing radiation. MRI scans are particularly valuable for examining soft tissues such as the brain, spinal cord, muscles, joints, and internal organs. They can provide information about the structure, function, and blood flow in these areas, aiding in the diagnosis and evaluation of various conditions, including tumors, injuries, neurological disorders, and cardiovascular diseases.

In this research, a novel framework of semi-supervised deep learning was proposed for COVID-19 image classification on chest x-ray images. Supervised learning based COVID-19 classification on x-ray datasets could provide useful information to medical staff for facilitating a diagnosis of COVID-19 in an effective and efficient manner.

Unfortunately, it relies on the availability of large amount of labeled medical images, which are not available in practice in the early outbreak of such global pandemic. Hence, a semi-supervised learning model based on ResNet that can utilize unlabeled images to enhance classification performance was proposed. There were two paths inthe model for reducing supervised cross entropy loss and unsupervised mean squared error loss, respectively. Then training was performed by jointly optimizing these two losses, which allowed the proposed scheme to take advantage of the information from both labeled and unlabeled images. Experimental results demonstrated that the proposed model could recognize COVID-19 lung pathology effectively by learning on very limited labeled images and substantial unlabeled images. For the future work, the plan is to extend the proposed model for other tasks such as COVID-19 image segmentation.

4.2 Finding for Calibrated Bagging Deep Learning for Image Semantic Segmentation: A Case Studyon Covid-19 Chest X-Ray Image Method.

4.2.1 Experimental Results

I validated the proposed method from two perspectives: comprehensive performance comparison between the baselines and the proposed method, and hyper-parameter examination.

4.2.1.1 Performance Comparison

Table 4.2 presents the performance comparison between the state-of-the-art individual models and the proposed method. It can be observed that these individual models could perform well on COVID-19 image segmentation regarding F1scores. and accuracy. Moreover, prediction uncertainties of most of them were promising with

respect to ECE and MCE. For these individual models, FCN32 ResNet50 outperformed other individual models with higher certainty. In addition, as one baseline, Ensemble(Majority Voting (MV)) performed better than other individual methods with highest prediction certainty by comparing F1score, ECE and MCE. It means that combining predictions of these individual models can effectively improve performance andprediction certainty. For the proposed method, Ensemble (Weighted Voting (ECE)) can perform better than the baselines including these individual models and ECE and MCE. For these individual models, FCN32 ResNet50 outperformed other individual models with higher certainty.

In addition, as one baseline, Ensemble(Majority Voting (MV)) performed better than other individual methods with highest prediction certainty by comparing F1score, ECE and MCE. It means that combining predictions of these individual models can effectively improve performance and prediction certainty. For the proposed method, Ensemble (Weighted Voting (ECE)) can perform betterthan the baselines including the individual models and the ensemble (Majority Voting(MV)) by comparing accuracy, recall, and F1score. Moreover, Ensemble (Weighted Voting (ECE)) was able to improve the prediction certainty. It means that using appropriate calibration errors as weights to implement weighted bagging deep learning can effectively improve prediction certainty as well as performance. In other words, it is an effective method to calibrate models by using appropriate calibration errors as weights to combine predictions. Furthermore, Ensemble (Majority Voting + ECE +MCE (MVEM)) obtained the optimal performance with highest prediction certainty. It indicated that ensemble strategy such as majority voting was effective to combine

predictions to further improve performance and prediction certainty. In addition to the performance comparison, an example of prediction visualization on semantic segmentation generated by the baselines and proposed models is shown in Fig. 4.4. When I examined the prediction visualization for these individual models, I observed that they missed some key components for detecting ROIs. Taking UNet as an example through comparing the predictions with ground truth, three key components circled are missed on subfigure (g). On the contrary, ensemble models such as MV, ECE, and MVEM perform better in that regard of predictions since they only miss one or two small components for detecting ROIs, where the proposed method including ECE and MVEM outperformed other baselines. It means that the proposed method can effectively improve recall on detecting ROIs by distributing contributions of prediction based on calibration errors such as ECE and MCE.

4.2.1.2 Hyper-Parameter Examination

Fine-tuning hyper-parameter for building deep learning models is an imperative step to obtain optimal performance. The process of building the proposed method involved various hyper-parameters. For example, for each individual DL model, the fine-tune learning rate, batch size, and epoch were fine-tuned to achieve optimal performance. Specially, for the proposed bagging deep learning, how many individual models involved is still an open challenge. Here, I individual models. Generally speaking, more individual models will enhance performance and improve prediction certainty regarding F1score and ECE. When we employed five individual models (Ensemble 5 (FCN32 RESNET50 + FCN32 + UNET + FCN32 MOBILENET + PSPNET)), the optimal performance and the highest prediction certainty regarding

values of accuracy were obtained, F1score, and ECE for Ensemble (Weighted Voting (ECE)) andEnsemble (Majority Voting + ECE + MCE (MVEM)), where the values of Recall and F1score were improved more significantly than other evaluation metrics. Additionally, Fig. 4.6 shows comparison of prediction visualization produced by the proposed methods built with different number of individual models. It was observed that more individual models involved in the proposed approach would reduce the size of missing components. Moreover, MVEM outperformed other ensemble methods, which means that majority voting based on more individual DL models can further enhance the performance of recognition of RIOs.

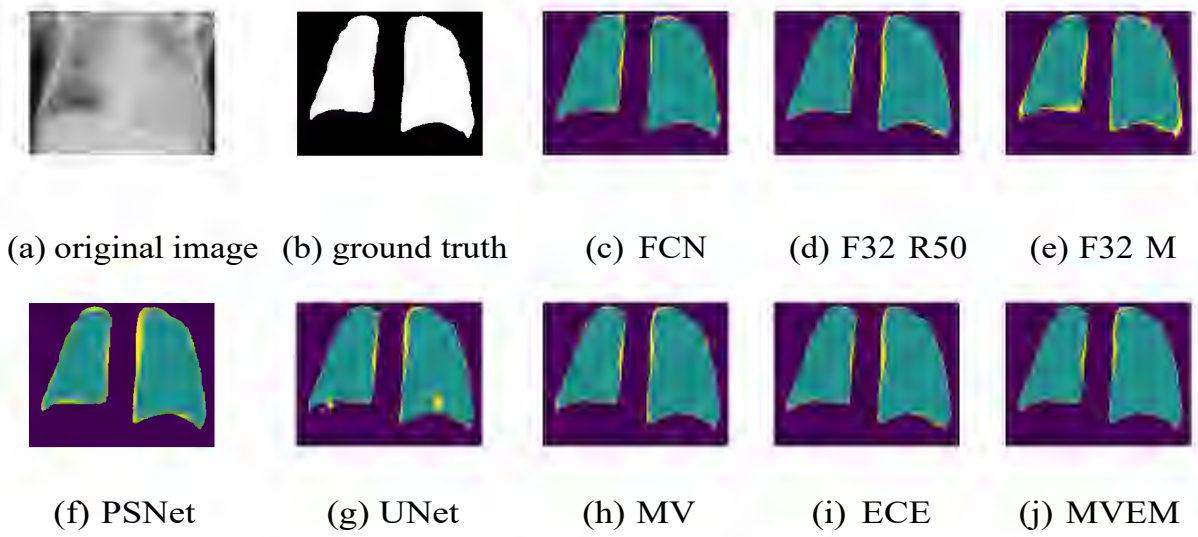


Fig. 4.5. An example of prediction visualization on semantic segmentation generated by the baselines and proposed models. F32 R50 and F32 M denotes FCN32 ResNet50 and FCN32 MobileNet while MV, ECE, and MVEM denotes Ensemble (Majority Voting (MV)), Ensemble (Weighted Voting (ECE)), and Ensemble (Majority Voting + ECE + MCE (MVEM)).

TABLE 4.4. COMPARING PERFORMANCE BETWEEN THE BASELINES AND THE PROPOSED METHOD USING ACCURACY (ACC), SENSITIVITY (SE), SPECIFICITY (SP), F1SCORE (F1), ECE AND MCE

DL	Acc (%)	Se (%)	Sp (%)	F1(%)	ECE (%)	MCE (%)
UNet	95.4±2.5	90.7±3.9	88.9±4.5	93.4±3.0	3.2±1.3	39.7±18.8
PSPNet	95.0±2.0	89.1±3.9	88.2±4.3	92.5±2.9	4.6±1.2	40.6±14.9
FCN32	95.8±2.4	92.3±4.5	91.0±5.0	94.0±3.5	2.5±2.1	37.6±19.1
FCN32 ResNet50 (F32 R50)	96.0±2.5	92.3±5.5	91.4±5.9	94.3±3.8	2.3±2.3	29.8±20.3
FCN32 MobileNet (F32 M)	95.2±2.3	91.0±4.7	90.1±5.6	93.1±3.3	4.1±1.6	38.2±19.9
Ensemble (Majority Voting (MV))	98.8±0.6	94.1±3.0	92.9±3.6	96.6±1.7	2.4±1.2	28.1±14.1
Ensemble (Weighted Voting (ECE))	99.1±0.5	95.4±2.9	94.3±2.9	97.1±1.5	2.3±1.2	24.7±12.4
Ensemble (Weighted Voting (MCE))	98.7±0.7	93.9±3.7	92.6±3.7	96.3±1.9	2.4±1.2	28.9±14.6
Ensemble (Majority Voting + ECE + MCE (MVEM))	99.2±0.4	97.7±2.3	95.4±2.3	98.4±0.8	2.1±1.1	20.1±10.1

In summary, in terms of observations mentioned above, the proposed method can effectively improve semantic segmentation, as well as reduce the prediction uncertainty through using the calibration error as weights of DL models to combine their predictions. Moreover, more individual DL models involved in the implementation of the proposed approach can further enhance the performance and

prediction certainty, which meets the intuition of majority voting for bagging deep learning. To some extent, it is an effective method to combine advantages of these individual DL models to improve the task performance without complex implementations.

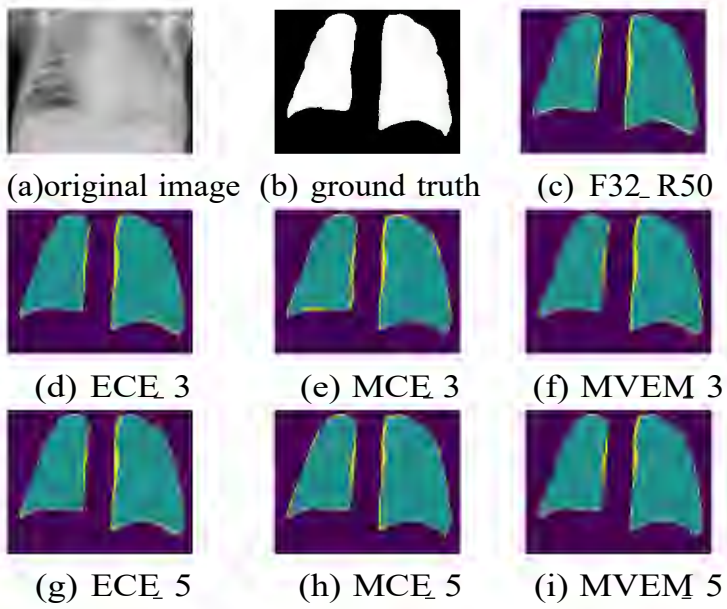


Fig. 4.6. Comparison of prediction visualization produced by the proposedmethods built with different number of individual models. The second row presents the predictions generated by three implementations of the proposed method with Ensemble 3 (FCN32 RESNET50 + FCN32 + UNET) including ECE 3 (Ensemble (Weighted Voting (ECE))), MCE 3 (Ensemble (Weighted Voting(MCE))), and MVEM 3 (Ensemble (Majority Voting + ECE + MCE (MVEM))). Similarly, The third row presents the predictions generated by five implementations of the proposed method with Ensemble 5 (FCN32 RESNET50 + FCN32 + UNET + FCN32 MOBILENET + PSPNET) including ECE 5, MCE 5, and MVEM 5.TABLE 4.5. COMPARING PERFORMANCE OF THE PROPOSED METHODS BUILT WITH DIFFERENT NUMBER OF INDIVIDUAL MODELS USING ACCURACY (ACC), SENSITIVITY (SE), SPECIFICITY (SP), F1SCORE (F1), ECE AND MCE

TABLE 4.5. COMPARING PERFORMANCE OF THE PROPOSED METHODS BUILT WITH DIFFERENT NUMBER OF INDIVIDUAL MODELS USING ACCURACY (ACC), SENSITIVITY (SE), SPECIFICITY (SP), F1SCORE (F1), ECE AND MCE

Ensemble 2 (FCN32 RESNET50 + FCN32)						
DL	Acc (%)	Se (%)	Sp (%)	F1 (%)	ECE(%)	MCE(%)
FCN32 ResNet50 (F32 R50)	95.8±2.1	92.3±3.9	91.0±4.5	94.0±3.0	2.5±1.3	37.6±18.8
Ensemble (Weighted Voting (ECE))	99.0±0.5	95.3±2.4	94.4±2.8	96.9±1.6	2.3±1.2	22.3±11.3
Ensemble (Weighted Voting (MCE))	98.8±0.6	93.8±3.1	93.7±3.2	96.4±1.8	2.5±1.3	25.1±12.6
Ensemble 3 (FCN32 RESNET50 + FCN32 + UNET)						
DL	Acc (%)	Se (%)	Sp (%)	F1 (%)	ECE(%)	MCE(%)
Ensemble (Majority Voting (MV))	98.7±0.7	93.9±3.0	94.1±3.0	96.1±2.0	2.7±1.4	31.1±15.6
Ensemble (Weighted Voting (ECE))	98.4±0.8	95.5±2.3	94.9±2.6	96.9±1.6	2.8±1.4	26.1±13.1
Ensemble (Weighted Voting (MCE))	98.3±0.9	93.1±3.5	93.0±3.5	96.0±2.0	2.8±1.4	31.2±15.6
Ensemble (Majority Voting + ECE + MCE (MVEM))	98.8±0.6	97.6±1.2	96.4±1.8	98.1±1.0	2.1±1.1	21.1±10.6
Ensemble 4 (FCN32 RESNET50 + FCN32 + UNET + FN32 MOBILENET)						
DL	Acc (%)	Se (%)	Sp (%)	F1 (%)	ECE(%)	MCE(%)
Ensemble (Weighted Voting (ECE))	98.3±0.9	95.0±2.5	94.6±2.7	96.5±1.8	2.7±1.4	24.3±13.5
Ensemble (Weighted Voting (MCE))	97.9±1.1	94.1±3.0	93.8±3.1	96.1±2.0	3.0±1.5	32.3±15.0
Ensemble 5 (FCN32 RESNET50 + FCN32 + UNET + FCN32 MOBILENET + PSPNET)						
DL	Acc (%)	Se (%)	Sp (%)	F1 (%)	ECE(%)	MCE(%)
Ensemble (Majority Voting (MV))	98.8±0.6	94.1±3.0	92.9±3.6	96.6±1.7	2.4±1.2	28.1±14.1
Ensemble (Weighted Voting (ECE))	99.1±0.5	95.4±2.9	94.3±2.9	97.1±1.5	2.3±1.2	24.7±12.4
Ensemble (Weighted Voting (MCE))	98.7±0.7	93.9±3.7	92.6±3.7	96.3±1.9	2.4±1.2	28.9±14.6
Ensemble (Majority Voting + ECE + MCE (MVEM))	99.2±0.4	97.7±2.3	95.4±2.3	98.4±0.8	2.1±1.1	20.1±10.1

4.2.2 Discussion

This research aimed to build a novel bagging learning method to implement COVID-19 semantic segmentation through combining bagging deep learning and model calibration. Semantic segmentation has achieved significant successes by developing deep learning models such as U-Net [2] and V-Net [137]. In the biomedical do- main, there have been numerous techniques for lung segmentation with different purposes [46, 138]. The U-Net is an effective technique for segmenting both lung regions and lung lesions in COVID applications [65]. The U-Net built with fully convolutional network [2] has a U-shape architecture with two symmetric paths: encoding path and decoding path. The layers at the same level in two paths are connected by the shortcut connections, which is to learn better visual semantics as well as detailed contexture. Zhou et al. [85] proposed the UNet++ that inserts a nested convolutional structure between the encoding and decoding path. In addition, Milletari et al. [137] built V-Net using the residual blocks as the basic convolutional block, and optimized the network by a Dice loss. Furthermore, Shan et al. [62] built VB-Net for more efficient segmentation by equipping the convolutional blocks with the so-called bottleneck blocks. Moreover, U-Net and its variants have been developed, achieving reasonable segmentation results in COVID-19 diagnosis [31]. In recent years, attention mechanisms can learn the most discriminant part of the features in deep learning models. Oktay et al. [139] proposed an Attention U-Netto capture fine structures in medical images, thereby suitable for segmenting lesionsand lung nodules in COVID-19 applications.

The main concern of such methods is associated with its high computation

complex and prior assumption on model weights. To reduce the computation complexity and enhance the scalability of Bayesian neural networks for data analysis on larger datasets, Hernńdez-Lobato et al. [140] proposed probabilistic back-propagation for learning Bayesian neural networks. Non-Bayesian-based methods develop various strategies such as model ensemble [141] and prior assumption on predictions [142]to estimate the prediction uncertainty, which is to reduce the cost of estimating the uncertainty. To reduce computation cost and training difficulty, Lakshminarayanan et al. [141] proposed deep ensemble that is simply to implement, trained in a parallel manner, requires less hyper-parameter tuning, and estimates high quality predictive uncertainty. However, it is very tricky to obtain the optimal number of individual models to build deep ensemble for various applications. Moreover, to reducethe cost of the memory usage and inference of Bayesian neural networks and deep ensembles, Liu (2020) et al. [143] proposed approaches to estimate uncertainty by building only one neural network with two steps: 1) measuring the distance between testing samples and training samples, and 2) implementing spectral-normalized neural Gaussian process (SNGP), that is, to improve the measurement of the distance by adding a weight normalization step during training and replacing the output layer with a Gaussian process. However, experimental results on dialog intent detection indicated that deep ensemble performed better than the proposed method on many evaluation metrics such as accuracy. Recently, Wilson et al. [144] systematically summarized Bayesian deep learning and claimed that deep ensemble could be treated as approximate Bayesian marginalization of model parameters. On the other side, theyalso claimed that Bayesian methods were not perfect regarding prior assumptions on model weights.

In this research, a novel bagging deep learning model was proposed for COVID-19 image segmentation on chest x-ray images. It combined the model calibration and traditional bagging learning to not only enhance the segmentation performance, but also improve the prediction certainty that is extremely important to high-risk applications in biomedical domain. The proposed method was validated on a large chest x-ray dataset that was associated with COVID-19. Experimental results demonstrated that the proposed model could recognize the lung region more effectively through comparing with state-of-the-art baselines. For the future work, the plan is to extend the proposed model for building an end-to-end model for both COVID-19 image classification and image segmentation.

4.3 Findings for Proformer-Based Ensemble Learning Gene Expression Prediction Method

4.3.1 Experimental setup

I proposed a novel ensemble of end-to-end Transformer encoder architectures that leveraged the self-attention mechanisms and could handle long-range correlations between the input-sequence items to predict the expression values from millions of DNA sequence. Specifically, the values of evaluation metrics including Spearman, PearsonR, Score Spearman, and Score PearsonR were employed as weight to implement the weighted sum during the prediction and average it to obtain the final predicted gene expression values. The key hyper-parameters of these individual models are shown in Table 4.6.The training losses consisted of the mean squared error between the expression values (y mse), and the reconstruction loss (y recon) which helps stabilize the training process especially with larges models, where five percent of the nucleotides were

randomly masked and predicted. The mean of the prediction of all heads for all predicted expression value from for each head was used as the final predicted expression value.

4.3.1 Experimental results

I validated the proposed method based on a comprehensive performance comparison between the baselines and the proposed method. Table 4.7 presents the performance comparison, where it includes benchmark performance of top three teams from DREAM Challenge 2022, variants of Proformers, and proposed weighted ensemble models. Specifically, these weighted ensemble models were built using weights obtained based on various evaluation metrics. It was observed that for the benchmarks, the winning team utosome.org outperformed Unlock DNA significantly. through examining variants of Proformer, the performance was similar across different models,which indicated that the model design did not impact the performance significantly

TABLE 4.6. HYPER-PARAMETERS OF BASELINE MODELS

Parameters	Model 1	Model 2	Model 3	Model 4
Attention Heads	4	8	8	8
Encoder Blocks	4	4	4	4
Expression Heads	NA	1	32	32
Batch size	512	512	512	512
Masking	NA	NA	5%	5%
Learning Rate	1e-3	1e-3	1e-3	1e-3
Epochs	20	20	20	20

For the proposed method, the weighted ensemble models performed better than variants of the proformer by comparing the PearsonR, Spearman, Score PearsonR, and Score Spearman values. It means that the ensemble model can effectively individual

model performance. In addition, employing different evaluation metrics to calculate weights led to different results, which illustrated that it should select evaluation metrics to obtain weights to achieve optimal performance. Finally, the proposed model outperformed the winning team regarding values of Score Spearman and Spearman, which further demonstrated the effectiveness of the proposed models.

TABLE 4.7. COMPARING PERFORMANCE BETWEEN THE PROPOSED ENSEMBLE MODELS AND BASELINES. THE BASELINES INCLUDE MODELS FROM TOP 3 TEAMS IN DREAM CHALLENGES2022 AND VARIANTS OF PROFORMER, THE PROPOSED MODELS ARE BUILT BASED ON PROFORMER FROM UNLOCK DNA.

Model	Score PearsonR	Score Spearman	PearsonR	Spearman
Top 3 teams in DREAM Challenges 2022				
utosome.org	0.818	0.854	0.972	0.976
BHI-dream challenge	0.791	0.845	0.962	0.971
Unlock DNA	0.766	0.823	0.957	0.967
Variants of Proformer				
Model1	0.766	0.819	0.918	0.961
Model2	0.765	0.817	0.921	0.964
Model3	0.781	0.827	0.926	0.965
Model4	0.765	0.810	0.929	0.967
Proposed Ensemble Models				
Weighted Ensemble (PearsonR)	0.781	0.833	0.931	0.977
Weighted Ensemble (Spearman)	0.793	0.857	0.939	0.981
Weighted Ensemble (Score PearsonR)	0.769	0.823	0.923	0.965
Weighted Ensemble (Score Spearman)	0.779	0.829	0.929	0.971

In addition to evaluating performance with various metrics, Fig. 4.7 compared differences between gene expression prediction levels and target gene expression levels, where the top subfigure contained the results with Model 4 while the bottom subfigure included predictions with the weighted ensemble model built with spearman values. It

randomly selected 50 genes in the prediction results for the comparison. Through comparing the difference between these two subfigures, it was obvious to observe that the weighted ensemble was able to enhance the predictions for different genes, for example, G23 and G27. It means that ensemble model was capable of using complementary advantages of multiple individual models to enhance the predictionsfor gene expression prediction.

4.3.2 Discussion

This study focused on predicting gene expression via developing novel deep learning models, which covered gene expression prediction and deep learning models. Although each cell in the human body contains a copy of an individual's DNA, not all genes are active or expressed in every type of cell at all times. The degree of a gene's expression is primarily governed by multiple input signals that are interpreted by the non-coding regulatory DNA sequences known as cis-regulatory logic.These sequences exert control over gene expression intensity by using transcription factors (TFs) that bind to regulatory sequences located throughout the DNA, including promoters that contain a wealth of information related to mRNA levels [75]. Gene regulation errors are frequently implicated in genetic disorders, as mutations that alter gene expression levels are a common cause. As such, comprehending the regulatory code would enable the development of cures for diseases and the management of protein production in biotechnology. Ultimately, a long-standing issue in regulatory genomics is associated with gene expression prediction from DNA sequence.

In the field of bioinformatics, gene expression prediction is an essential task. Gene expression involves using the information stored in genes to create a functional gene

product. By connecting the expression of genes of interest to a biological process or phenotype, researchers can gain insights into gene function, biological pathways, and the genes responsible for regulating development, cell behavior, and signaling [75]. Models that use DNA sequences to estimate gene expression offer the potential to enhance understanding of transcriptional regulation and the impact of non-coding genetic variants associated with diseases and traits. These models can take two different approaches: mechanistic models [154], which aim to simulate the underlying biology directly, and artificial intelligence (AI) models [155], which do not necessarily use existing biological knowledge but instead learn to map input cell states to output expression levels. Predicting gene expression involves estimating the expression levels of many genes simultaneously, which has led to the application of multi-task learning, transfer learning, and deep learning methods [156, 157].

Deep learning is a powerful AI technique that has the ability to automatically extract features by efficiently exploring the feature space and identifying nonlinear transformations of weighted averages of those features [158]. One example of deep learning application is deepChrome [159], which employs a convolutional neural network (CNN) to predict gene expression levels based on five types of histone mark ChIP-seq signals aggregated in a 10, 000bp region surrounding the transcription start site (TSS) of each gene. To visualize feature importance within the gene's local neighborhood, deepChrome introduced attention layers [159]. Another deep learning system, deepTrio [160], uses a mask multiple parallel CNN to predict protein-protein interaction. Temporal Convolutional Networks (TCN), a type of stacked neural network where each hidden layer has the same length as the input layer, was designed to

ensure that predictions for a target time point are dependent on all the previous time points' information [154]. Another CNN model with six convolutional layers, three residual connections, and additional features like batch normalization, dropouts for regularization, and max pooling operation after the penultimate convolutional layer [158] was implemented to improve generalization and reduce model size while learning feature representations of promoter sequences. In the Dream Challenges, various CNN implementations were utilized in the benchmark methods [146] to learn feature representations of the nucleotide combinations of the entire target promoter sequence and predict expression profiles.

CNN-based models have a limitation in information flow between distal elements, restricting them to consider only sequence elements up to 20kb away from the transcription start site (TSS) [161]. To overcome this limitation, Graph Convolution Network (GCN) has been developed as a powerful tool for generalizing the traditional convolution to graphs by propagating information of neighboring nodes for each central node [162]. Additionally, attention mechanisms have been integrated into many models to assign weights to different parts of input data, which can im- prove prediction performance. In this study, attention mechanism in transformer encoder was used to build novel ensemble models to further improve performance inlearning feature representations for gene expression prediction.

In this experiment, a novel ensemble deep learning model was proposed for gene expression prediction. It combined transformer and traditional bagging learning to not only enhance the prediction performance, but also validate a simple and effectivestrategy to enhance benchmarks based on large deep learning models. Experimental results

illustrated that the proposed model could predict gene expressions effectively through comparing with state-of-the-art baselines. For the future work, we plan to further enhance the propose

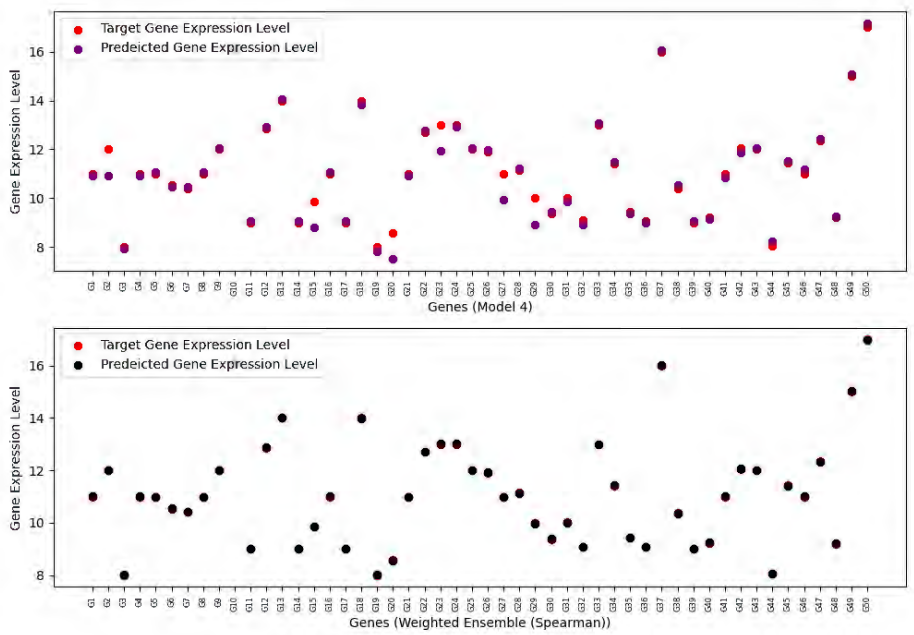


Fig. 4.7. Differences between gene expression prediction levels and target gene expression levels. x-axis denotes different genes while y-axis denotes gene expression levels.

CHAPTER 5

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

The goal of this study was to apply deep learning techniques to biomedical data analysis. With a focus on analyzing biomedical images, a generalized model for the detection of covid-19 on radiological images was proposed. Deep learning methods are applied to the processing and analysis of chest x-ray images as potential solution to the early detection and diagnosis of Covid-19. Many research efforts have been focused on detecting the covid-19 patterns on chest images but the unavailability of labeled chest images and ensuring the reliability of the AI modelshas been a continuous issue and this research ~~has~~ introduced two approaches to address these limitations. A semi-supervised learning model based on ResNet that can utilize unlabeled images to enhance classification performance was proposed. There were two paths in the model for reducing supervised cross entropy loss and unsupervised mean squared error loss, respectively. Then training was performed by jointly optimizing these two losses, which allowed the proposed scheme to take advantage of the information from both labeled and unlabeled images. Experimental results demonstrated that the proposed model could recognize COVID-19 lung pathology effectivelyby learning on very limited labeled images and substantial unlabeled images. A second approach that involved bagging deep learning model was proposed for COVID-19 image segmentation on chest x-ray images. It combined the model calibration and traditional bagging learning to not only enhance the segmentation performance, but also improve the prediction certainty that is extremely important to high-risk applications in biomedical domain.

I validated the proposed method on a large chest x-ray dataset that was associated with COVID-19. Experimental results demonstrated that the proposed model could recognize the lung region more effectively by comparing with state-of-the-art baselines. Another focus of this work was to explore building high-performance deep learning models to enhance gene expression prediction using high-throughput measurements of cis-regulatory activity in yeast through randomly generating millions of promoters. In this research, a novel ensemble deep learning model was proposed for gene expression prediction, which combines the transformer architecture and traditional bagging learning. This model not only enhanced prediction performance but also introduced a simple and effective strategy to improve benchmarks based on large deep learning models. Experimental results demonstrated the model's effectiveness in predicting gene expressions compared to state-of-the-art baselines.

5.2 Recommendations

In future work, it is recommended that the proposed ensemble deep learning model for gene expression prediction, which combines the transformer architecture and traditional bagging learning, be further enhanced by incorporating novel data augmentation techniques, such as GPT-based data augmentation. The GPT model is based on the transformer architecture used in this research for gene expression prediction, where it has a stack of transformer layers, each layer has a multi-head self-attention mechanism and a feed-forward neural network. The self-attention mechanism allows the model to capture relationships between tokens in a sequence, while the feed-forward neural network helps in incorporating non-linear transformations.

In addition, few-shot learning techniques were proposed to improve the model's

ability to generalize and make accurate predictions when given only a limited amount of labeled data. By leveraging the language processing capabilities of ChatGPT, the GPT model can be trained to learn from a small number of labeled examples and effectively apply that knowledge to unseen gene expression data. This approach has the potential to enhance the model's performance and expand its applicability in scenarios where labeled data is scarce or limited.

REFERENCES

[1] F. Rousseau, L. Drumetz, and R. Fablet, "Residual Networks as Flows of Ddiffeomorphisms," *Journal of Mathematical Imaging and Vision*, vol. 62, 04 2020.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation, arxiv. org," in *Computer Vision and Pattern Recognition*, 2015.

[3] J. Lovén, D. A. Orlando, A. A. Sigova, C. Y. Lin, P. B. Rahl, C. B. Burge, D. L. Levens, T. I. Lee, and R. A. Young, "Revisiting global gene expression analysis," *Cell*, vol. 151, no. 3, pp. 476–482, 2012.

[4] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88,2017.

[5] "Proformer," *https: // github. com/ gongx030/ dream_ PGE*, 2022.

[6] L. C. Greig, M. B. Woodworth, M. J. Galazo, H. Padmanabhan, and J. D. Macklis, "Molecular logic of neocortical projection neuron specification, development and diversity," *Nature Reviews Neuroscience*, vol. 14, no. 11, pp. 755–769, 2013.

[7] J. Luo, M. Wu, D. Gopukumar, and Y. Zhao, "Big data application in biomedical research and health care: a literature review," *Biomedical informatics insights*, vol. 8, pp. BII–S31559, 2016.

[8] R. F. Woolson and W. R. Clarke, *Statistical methods for the analysis of biomedical data*. John Wiley & Sons, 2011.

[9] M. Jeon, H. Jeong, K. Lee, and J. Yim, "The acute effect of smoking a single cigarette on vascular status, spo2, and stress level," *Medical science monitor: international medical journal of experimental and clinical research*, vol. 20,p. 601, 2014.

[10] C. Miranda-Fuentes, L. J. Chirosa-R´ıos, I. M. Guisado-Requena, P. Delgado-Floody, and D. Jerez-Mayorga, "Changes in muscle oxygen saturation mea-

sured using wireless near-infrared spectroscopy in resistance training: A systematic review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 8, p. 4293, 2021.

[11] A. Subramaniana, P. Tamayoa, V. K. Moothaa, S. Mukherjeed, B. L. Eberta, M. A. Gillettea, A. Paulovichg, S. L. Pomeroyh, T. R. Goluba, E. S. Landera,*et al.*, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *PNAS*, vol. 102, no. 43, pp. 15545–15550, 2005.

[12] U. S. Shanthamallu, A. Spanias, C. Tepedelenlioglu, and M. Stanley, "A briefsurvey of machine learning methods and their sensor and iot applications,"in *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pp. 1–8, IEEE, 2017.

[13] S. Suthaharan, "Support vector machine," in *Machine learning models and algorithms for big data classification*, pp. 207–235, Springer, 2016.

[14] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, p. 130, 2015.

[15] D. Greene, P. Cunningham, and R. Mayer, "Unsupervised learning and clustering," in *Machine learning techniques for multimedia*, pp. 51–90, Springer, 2008.

[16] S. Na, L. Xumin, and G. Yong, "Research on k-means clustering algorithm: An improved k-means clustering algorithm," in *2010 Third International Symposium on intelligent information technology and security informatics*, pp. 63–67,Ieee, 2010.

[17] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, pp. 1–21, 2021.

[18] S. Sah, "Machine learning: a review of learning types," 2020.

[19] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[20] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," 2020.

[21] L. Li, L. Han, M. Ding, H. Cao, and H. Hu, "A deep learning semantic template matching framework for remote sensing image registration," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 181, pp. 205–217, 2021.

[22] D. Joo, E. Yi, S. Baek, and J. Kim, "Linearly replaceable filters for deep network channel pruning," in *Proceedings of the AAAI Conference on ArtificialIntelligence*, vol. 35, pp. 8021–8029, 2021.

[23] W. L. Alyoubi, M. F. Abulkhair, and W. M. Shalash, "Diabetic retinopathy fundus image classification and lesions localization system using deep learning," *Sensors*, vol. 21, no. 11, p. 3704, 2021.

[24] L.-s. Wei, Q. Gan, and T. Ji, "Skin disease recognition method based on image color and texture features," *Computational and mathematical methods inmedicine*, vol. 2018, 2018.

[25] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact cnns," *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–17, 2019.

[26] R. Kundu, R. Das, Z. W. Geem, G.-T. Han, and R. Sarkar, "Pneumonia detection in chest x-ray images using an ensemble of deep learning models," *Plos one*, vol. 16, no. 9, p. e0256630, 2021.

[27] R. Yi, L. Tang, Y. Tian, J. Liu, and Z. Wu, "Identification and classificationof pneumonia disease using a deep learning-based intelligent computational framework," *Neural Computing and Applications*, pp. 1–14, 2021.

[28] T. S. Sheikh, Y. Lee, and M. Cho, "Histopathological classification of breast cancer images using a multi-scale input and multi-feature network," *Cancers*, vol. 12, no. 8, p. 2031, 2020.

[29] M. M. Althobaiti, A. Almulihi, A. A. Ashour, R. F. Mansour, and D. Gupta, "Design of optimal deep learning-based pancreatic tumor and nontumor classification model using computed tomography scans," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[30] C. Zhao, R. Shuai, L. Ma, W. Liu, and M. Wu, "Segmentation of skin lesions image based on u-net++," *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 8691–8717, 2022.

[31] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: a review," *Frontiers in CardiovascularMedicine*, vol. 7, p. 25, 2020.

[32] D. Müller and F. Kramer, "Miscnn: a framework for medical image segmentation with convolutional neural networks and deep learning," *BMC medical imaging*, vol. 21, no. 1, pp. 1–11, 2021.

[33] A. Yang, W. Zhang, J. Wang, K. Yang, Y. Han, and L. Zhang, "Review on the application of machine learning algorithms in the sequence data mining of dna," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 1032, 2020.

[34] Z. Wu and C. Shen, "Hengel avd," *Bridging category-level and instance-level semantic image segmentation. arXiv preprint arXiv*, vol. 160506885, 2016.

[35] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for se- mantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[36] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolu- tional neural network," in *2017 international conference on engineering and technology (ICET)*, pp. 1–6, IEEE, 2017.

[37] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, 2017.

[38] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer,"Densenet: Implementing efficient convnet descriptor pyramids," *arXiv preprint arXiv:1404.1869*, 2014.

[39] J. Chen, Q. Zou, and J. Li, "Deepm6aseq-el: prediction of human n6-methyladenosine (m 6 a) sites with lstm and ensemble learning," *Frontiers of Computer Science*, vol. 16, pp. 1–7, 2022.

[40] Y. Hu, L. Zhao, Z. Li, X. Dong, T. Xu, and Y. Zhao, "Classifying the multi- omics data of gastric cancer using a deep feature selection method,"

Expert Systems with Applications, vol. 200, p. 116813, 2022.

[41] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of dna-and rna-binding proteins by deep learning," *Nature biotechnology*, vol. 33, no. 8, pp. 831–838, 2015.

[42] Y. Zeng, Z. Wei, W. Yu, R. Yin, Y. Yuan, B. Li, Z. Tang, Y. Lu, and Y. Yang, "Spatial transcriptomics prediction from histology jointly through transformerand graph neural networks," *Briefings in Bioinformatics*, vol. 23, no. 5, 2022.

[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[44] V. Van Asch, "Macro-and micro-averaged evaluation measures [[basic draft]]," *Belgium: CLiPS*, vol. 49, 2013.

[45] M. D. McKay, "Evaluating prediction uncertainty," tech. rep., Nuclear Regulatory Commission, 1995.

[46] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, pp. 424–432, Springer, 2016.

[47] K. Preston, *Digital processing of biomedical images*. Springer Science & Busi-ness Media, 2012.

[48] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, et al., "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface*, vol. 15, no. 141, p. 20170387, 2018.

[49] L. Liang, M. Liu, C. Martin, and W. Sun, "A deep learning approach to estimate stress distribution: a fast and accurate surrogate of finite-element analysis," *Journal of The Royal Society Interface*, vol. 15, no. 138, p. 20170844, 2018.

[50] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov, "Applications of deep learning in biomedicine," *Molecular pharmaceutics*, vol. 13, no. 5, pp.

1445– 1454, 2016.

[51] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Molecular Systems Biology*, vol. 12, no. 7, p. 878, 2016.

[52] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in Bioinformatics*, vol. 18, no. 5, pp. 851–869, 2017.

[53] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," *Journal of Imaging*, vol. 6, no. 6, p. 52, 2020.

[54] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, and L. Xia, "Serial quantitative chest ct assessment of covid-19: Deep-learning approach," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200075, 2020.

[55] S. Bharati, P. Podder, M. Mondal, and V. Prasath, "Medical imaging with deep learning for covid-19 diagnosis: a comprehensive review," *arXiv preprint arXiv:2107.09602*, 2021.

J. Yang, Y. Zheng, X. Gou, K. Pu, Z. Chen, Q. Guo, R. Ji, H. Wang, Y. Wang, and Y. Zhou, "Prevalence of comorbidities and its effects in patients infected with sars-cov-2: a systematic review and meta-analysis," *International Journalof Infectious Diseases*, vol. 94, pp. 91–95, 2020.

[56] T. Liu, E. Siegel, and D. Shen, "Deep learning and medical image analysis forcovid-19 diagnosis and prediction," *Annual Review of Biomedical Engineering*,vol. 24, pp. 179–201, 2022.

[57] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *arXiv preprint arXiv:2003.09871*, 2020.

[58] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection fromx-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635–640, 2020.

[59] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid-19 from chest ct using weak label,"

medRxiv, 2020.

[60] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," *arXiv preprint arXiv:2003.05037*,2020.

[61] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, and Y. Shi, "Lung infection quantification of covid-19 in ct images with deep learning," *arXiv preprint arXiv:2003.04655*, 2020.

[62] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, *et al.*, "Ai-assisted ct imaging analysis for covid-19 screening: Build- ing and deploying a medical ai system in four weeks," *medRxiv*, 2020.

[63] L. Tang, X. Zhang, Y. Wang, and X. Zeng, "Severe covid-19 pneumonia: Assessing inflammation burden with volume-rendered chest ct," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200044, 2020.

[64] Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, and H. Shi, "Longitudinal assessment of covid-19 using a deep learning–based quantitative ct pipeline: Illustration of two cases," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200082, 2020.

[65] G. Gaál, B. Maga, and A. Lukács, "Attention u-net based adversarial architectures for chest x-ray lung segmentation," *arXiv preprint arXiv:2003.10304*, 2020.

[66] J. Weston, F. Ratle, and R. Collobert, "Deep learning via semi-supervised embedding," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 1168–1175, 2008.

[67] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Advances in neural information processing systems*, pp. 3581–3589, 2014.

[68] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *Advances in Neural Information Processing Systems*, pp. 3546–3554, 2015.

[69] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," *arXiv preprint arXiv:1610.02242*, 2016.

[70] L. Nwosu, X. Li, L. Qian, S. Kim, and X. Dong, "Calibrated bagging deep learning for image semantic segmentation: A case study on covid-19 chest x-ray image," *PLoS One*, vol. 17, no. 11, p. e0276250, 2022.

[71] L. Nwosu, X. Li, L. Qian, S. Kim, and X. Dong, "Semi-supervised learning for covid-19 image classification via resnet," *arXiv preprint arXiv:2103.06140*, 2021.

[72] X. Zeng, S. Zhu, X. Liu, Y. Zhou, R. Nussinov, and F. Cheng, "deepdr: a network-based deep learning approach to in silico drug repositioning," *Bioinformatics*, vol. 35, no. 24, pp. 5191–5198, 2019.

[73] O. A. Sarumi and C. K. Leung, "Adaptive machine learning algorithm and analytics of big genomic data for gene prediction," *Tracking and preventing diseases with artificial intelligence*, pp. 103–123, 2022.

[74] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Studying gene expression and function," in *Molecular Biology of the Cell. 4th edition*,Garland Science, 2002.

[75] L. M. Blaabjerg, M. M. Kassem, L. L. Good, N. Jonsson, M. Cagiada, K. E. Johansson, W. Boomsma, A. Stein, and K. Lindorff-Larsen, "Rapid protein stability prediction using deep learning representations," *Elife*, vol. 12, p. e82593,2023.

[76] P. Moeskops, J. M. Wolterink, B. H. Van Der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum, "Deep learning for multi-task medical image seg- mentation in multiple modalities," in *Medical Image Computing and Computer- Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pp. 478–486, Springer, 2016.

[77] C. P. Langlotz, B. Allen, B. J. Erickson, J. Kalpathy-Cramer, K. Bigelow, T. S. Cook, A. E. Flanders, M. P. Lungren, D. S. Mendelson, J. D. Rudie, *et al.*, "A roadmap for foundational research on artificial intelligence in medical imaging:from the 2018 nih/rsna/acr/the academy workshop," *Radiology*,

vol. 291, no. 3, pp. 781–791, 2019.

[78] "NY Times: Germany got testing right, what can we learn?" Accessed: 2020-04-29.

[79] "CDC: Testing for COVID-19." urlhttps://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html. Accessed: 2020-04-19.

[80] J. Bullock, K. H. Pham, C. S. N. Lam, M. Luengo-Oroz, et al., "Mapping the landscape of artificial intelligence applications against covid-19," *arXiv preprintarXiv:2003.11336*, 2020.

[81] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, seg- mentation and diagnosis for covid-19," *IEEE Reviews in Biomedical Engineering*,2020.

[82] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *MedRxiv*, 2020.

[83] C. Butt, J. Gill, D. Chun, and B. A. Babu, "Deep learning system to screen coronavirus disease 2019 pneumonia," *Applied Intelligence*, p. 1, 2020.

[84] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nestedu-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11, Springer, 2018.

[85] C. Jin, W. Chen, Y. Cao, Z. Xu, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, and J. Feng, "Development and evaluation of an ai system for covid-19 diagnosis," *medRxiv*, 2020.

[86] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[87] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection," *arXiv preprint arXiv:2003.10769*, 2020.

[88] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection," *arXiv*

preprint arXiv:2003.11597, 2020.

[89] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849*, 2020.

[90] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," *arXiv preprint arXiv:2003.12338*, 2020.

[91] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, vol. 3, 2013.

[92] J. Zhou, B. Jing, and Z. Wang, "Soda: Detecting covid-19 in chest x- rays with semi-supervised open set domain adaptation," *arXiv preprint arXiv:2005.11003*, 2020.

[93] S. Calderon-Ramirez, A. Moemeni, D. Elizondo, S. Colreavy-Donnelly, L. F. Chavarria-Estrada, M. A. Molina-Cabello, *et al.*, "Correcting data imbalance for semi-supervised covid-19 detection using x-ray chest images," *arXiv preprint arXiv:2008.08496*, 2020.

[94] A. Paticchio, T. Scarlatti, M. Mattheakis, P. Protopapas, and M. Brambilla, "Semi-supervised neural networks solve an inverse problem for modeling covid-19 spread," *arXiv preprint arXiv:2010.05074*, 2020.

[95] J. Ma, Z. Nie, C. Wang, G. Dong, Q. Zhu, J. He, L. Gui, and X. Yang, "Active contour regularized semi-supervised learning for covid-19 ct infection segmentation with limited annotations," *Physics in Medicine & Biology*, 2020.

[96] D. Yang, Z. Xu, W. Li, A. Myronenko, H. R. Roth, S. Harmon, S. Xu, B. Turk- bey, E. Turkbey, X. Wang, et al., "Federated semi-supervised learning for covidregion segmentation in chest ct using multi-national data from china, italy, japan," *arXiv preprint arXiv:2011.11750*, 2020.

[97] A. D. Berenguer, H. Sahli, B. Joukovsky, M. Kvasnytsia, I. Dirks, M. Alioscha-Perez, N. Deligiannis, P. Gonidakis, S. A. Sánchez, R. Brahimetaj, et *al.*, "Explainable-by-design semi-supervised representation learning for covid-19 diagnosis from ct imaging," *arXiv preprint*

arXiv:2011.11719, 2020.

[98] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *European Conference on Computer Vision*, pp. 94–108, Springer, 2014.

[99] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[100] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[101] Q. Liu, L. Yu, L. Luo, Q. Dou, and P. A. Heng, "Semi-supervised medical image classification with relation-driven self-ensembling model," *IEEE Transactions on Medical Imaging*, 2020.

[102] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,"in *Advances in neural information processing systems*, pp. 1195–1204, 2017.

[103] F. Zhu, H. Li, W. Ouyang, N. Yu, and X. Wang, "Learning spatial regularization with image-level supervisions for multi-label image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5513–5522, 2017.

[104] H. Borgli, V. Thambawita, P. H. Smedsrud, S. Hicks, D. Jha, S. L. Eskeland, K. R. Randel, K. Pogorelov, M. Lux, D. T. D. Nguyen, *et al.*, "Hyperkvasir,a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Scientific Data*, vol. 7, no. 1, pp. 1–14, 2020.

[105] Y. Yang, "A study of thresholding strategies for text categorization," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 137–145, 2001.

[106] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data mining and knowledge discovery handbook*, pp. 875–886, Springer, 2009.

[107] X. Wang, X. Song, Y. Guan, B. Li, and J. Han, "Comprehensive named entity recognition on cord-19 with distant or weak supervision," *arXiv preprint arXiv: 2003.12218,* 2020.

[108] Ş. Öztürk, U. Özkaya, and M. Barstuğan, "Classification of coronavirus (covid- 19) from x-ray and ct images using shrunken features," *International Journal of Imaging Systems and Technology*, vol. 31, no. 1, pp. 5–15, 2021.

[109] X. Zhang, S. Lu, S.-H. Wang, X. Yu, S.-J. Wang, L. Yao, Y. Pan, and Y.-D. Zhang, "Diagnosis of covid-19 pneumonia via a novel deep learning architec- ture," *Journal of Computer Science and Technology*, vol. 1, 2021.

[110] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R. L. Hes- keth, L. Yang, *et al.*, "Time course of lung changes on chest ct during recov-ery from 2019 novel coronavirus (covid-19) pneumonia," *Radiology*, p. 200370,2020.

[111] H. Y. F. Wong, H. Y. S. Lam, A. H.-T. Fong, S. T. Leung, T. W.-Y. Chin, C. S. Y. Lo, M. M.-S. Lui, J. C. Y. Lee, K. W.-H. Chiu, T. Chung, et al., "Frequency and distribution of chest radiographic findings in covid-19 positive patients," *Radiology*, p. 201160, 2020.

[112] M. Kumar, S. Gupta, K. Kumar, and M. Sachdeva, "Spreading of covid-19 in India, Italy, Japan, Spain, UK, US: A prediction using arima and lstm model," *Digital Government: Research and Practice*, vol. 1, no. 4, pp. 1–9, 2020.

[113] T. Liang *et al.*, "Handbook of covid-19 prevention and treatment," *Zhejiang: Zhejiang University School of Medicine*, 2020.

[114] Y. Wang, M. Hu, Q. Li, X.-P. Zhang, G. Zhai, and N. Yao, "Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with covid-19 in an accurate and unobtrusive manner," *arXiv preprint arXiv:2002.05534*, 2020

[115] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055*, 2020.

[116] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

[117] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern*

Recognition, pp. 1251–1258, 2017.

[118] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[119] P. K. Sethy, S. K. Behera, P. K. Ratha, and P. Biswas, "Detection of coronavirus disease (covid-19) based on deep features and support vector machine," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 5, no. 4, 2020.

[120] I. Castiglioni, D. Ippolito, M. Interlenghi, C. B. Monti, C. Salvatore, S. Schi- affino, A. Polidori, D. Gandola, C. Messa, and F. Sardanelli, "Artificial intelligence applied on chest x-ray can aid in the diagnosis of covid-19 infection: a first experience from lombardy, italy," *MedRxiv*, 2020.

[121] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 1207–1220, 2021.

[122] H. Y. Paul, T. K. Kim, and C. T. Lin, "Generalizability of deep learning tuberculosis classifier to covid-19 chest radiographs: new tricks for an old algorithm?" *Journal of Thoracic Imaging*, vol. 35, no. 4, pp. W102–W104, 2020.

[123] L. O. Teixeira, R. M. Pereira, D. Bertolini, L. S. Oliveira, L. Nanni, G. D. Cavalcanti, and Y. M. Costa, "Impact of lung segmentation on the diagnosis and explanation of covid-19 in chest x-ray images," *Sensors*, vol. 21, no. 21, p. 7116, 2021.

[124] S. Chakraborty, A. K. Saha, S. Nama, and S. Debnath, "Covid-19 x-ray im- age segmentation by modified whale optimization algorithm with population reduction," *Computers in Biology and Medicine*, vol. 139, p. 104984, 2021.

[125] A. Bhattacharyya, D. Bhaik, S. Kumar, P. Thakur, R. Sharma, and R. B. Pachori, "A deep learning based approach for automatic detection of covid- 19 cases using chest x-ray images," *Biomedical Signal Processing and Control*,

vol. 71, p. 103182, 2022.

[126] M. Ganaie, M. Hu, *et al.*, "Ensemble deep learning: A review," *arXiv preprint arXiv:2104.02395*, 2021.

[127] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International Conference on Machine Learning*, pp. 1321–1330, PMLR, 2017.

[128] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[129] A. H. Murphy and E. S. Epstein, "Verification of probabilistic predictions: Abrief review," *Journal of Applied Meteorology and Climatology*, 6, no.5, pp. 748–755, 1967

[130] C. S. Crowson, E. J. Atkinson, and T. M. Therneau, "Assessing calibration of prognostic risk scores," *Statistical methods in medical research*, vol. 25, no. 4, pp. 1692–1706, 2016.

[131] X. Jiang, M. Osl, J. Kim, and L. Ohno-Machado, "Calibrating predictive model estimates to support personalized medicine," *Journal of the American Medical Informatics Association*, vol. 19, no. 2, pp. 263–274, 2012.

[132] B. Sayin, E. Krivosheev, J. Ram´ırez, F. Casati, E. Taran, V. Malanina, and J. Yang, "Crowd-powered hybrid classification services: Calibration is all you need," in *2021 IEEE International Conference on Web Services (ICWS)*, pp. 42–50, IEEE, 2021.

[133] M. P. Naeini, G. Cooper, and M. Hauskrecht, "Obtaining well calibrated probabilities using bayesian binning," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[134] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[135] A. Saood and I. Hatem, "Covid-19 lung ct image segmentation using deep learning methods: U-net versus segnet," *BMC Medical Imaging*, vol. 21, no. 1, pp. 1–10, 2021.

[136] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neuralnetworks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, IEEE, 2016.

[137] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, et al., "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.

[138] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[139] J. M. Hernández-Lobato and R. Adams, "Probabilistic backpropagation for scalable learning of bayesian neural networks," in *International Conference on Machine Learning*, pp. 1861–1869, PMLR, 2015.

[140] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *arXiv preprint arXiv:1612.01474*, 2016.

[141] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," *arXiv preprint arXiv:1806.01768*, 2018.

[142] J. Z. Liu, Z. Lin, S. Padhy, D. Tran, T. Bedrax-Weiss, and B. Lakshminarayanan, "Simple and principled uncertainty estimation with deterministic deep learning via distance awareness," *arXiv preprint arXiv:2006.10108*, 2020.

[143] A. G. Wilson, "The case for bayesian deep learning," *arXiv preprint arXiv:2001.10995*, 2020.

[144] "Dream challenge," *https://sagebionetworks.org/research-projects/dream-challenges-powered-by-sage-bionetworks/*, 2007.

[145] "Dream challenge 2022," *https://www.synapse.org/\#\!Synapse:syn28469146/wiki/617075*, 2022.

[146] R. Xie, J. Wen, A. Quitadamo, J. Cheng, and X. Shi, "A deep auto-encoder model for gene expression prediction," *BMC genomics*, vol. 18, pp. 39–49, 2017.

[147] J. Zhou, C. L. Theesfeld, K. Yao, K. M. Chen, A. K. Wong, and O. G. Troyanskaya, "Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk," *Nature Genetics*, vol. 50, no. 8, pp. 1171–1179, 2018.

[148] D. R. Kelley, "Cross-species regulatory sequence activity prediction," *PLoS computational biology*, vol. 16, no. 7, p. e1008050, 2020.

[149] V. Agarwal and J. Shendure, "Predicting mrna abundance directly from genomic sequence using deep convolutional neural networks," *Cell Reports*, vol. 31, no. 7, p. 107663, 2020.

[150] D. R. Kelley, Y. A. Reshef, M. Bileschi, D. Belanger, C. Y. McLean, and J. Snoek, "Sequential regulatory activity prediction across chromosomes with convolutional neural networks," *Genome Research*, vol. 28, no. 5, pp. 739–750,2018.

[151] E. van den Heuvel and Z. Zhan, "Myths about linear and monotonic associations: pearson'sr, spearman's ρ, and kendall's τ," *The American Statistician*, vol. 76, no. 1, pp. 44–52, 2022.

[152] J. F. Costa, "Weighted correlation," in *international encyclopedia of statistical science*, pp. 1653–1655, Springer, 2011.

[153] V. Pipoli, M. Cappelli, A. Palladini, C. Peluso, M. Lovino, and E. Ficarra, "Predicting gene expression levels from dna sequences and post-transcriptional information with transformers," *Computer Methods and Programs in Biomedicine*, vol. 225, p. 107035, 2022.

[154] N. Al Taweraqi and R. D. King, "Improved prediction of gene expression through integrating cell signalling models with machine learning," *BMC bioinformatics*, vol. 23, no. 1, p. 323, 2022.

[155] Ž. Avsec, V. Agarwal, D. Visentin, J. R. Ledsam, A. Grabska-Barwinska, K. R. Taylor, Y. Assael, J. Jumper, P. Kohli, and D. R. Kelley, "Effective gene expression prediction from sequence by integrating long-range interactions," *Nature Methods*, vol. 18, no. 10, pp. 1196–1203, 2021.

[156] R. Li, H. Xu, and E. A. Mukamel, "Epiattend: A transformer model of gene regulation combining single cell epigenomes with dna sequence," in

NeurIPS 2022 Workshop on Learning Meaningful Representations of Life.

[157] E. Alhenawi, R. Al-Sayyed, A. Hudaib, and S. Mirjalili, "Feature selection methods on gene expression microarray data for cancer classification: A systematic review," *Computers in Biology and Medicine*, vol. 140, p. 105051, 2022.

[158] F. Frasca, M. Matteucci, M. Leone, M. J. Morelli, and M. Masseroli, "Accurate and highly interpretable prediction of gene expression from histone modifications," *BMC bioinformatics*, vol. 23, no. 1, pp. 1–17, 2022.

[159] X. Hu, C. Feng, Y. Zhou, A. Harrison, and M. Chen, "Deeptrio: A ternary prediction system for protein–protein interaction using mask multiple parallel convolutional neural networks," *Bioinformatics*, vol. 38, no. 3, pp. 694–702, 2022.

[160] P. Zheng, G. Zhang, Y. Liu, and G. Huang, "Multiscale-cnn-4mcpred: a multi-scale cnn and adaptive embedding-based method for mouse genome dna n4-methylcytosine prediction," *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–15, 202

[161] J. Bigness, X. Loinaz, S. Patel, E. Larschan, and R. Singh, "Integrating long-range regulatory interactions to predict gene expression using graph convolutional networks," *Journal of Computational Biology*, vol. 29, no. 5, pp. 409–424, 2022.

CURRICULUM VITA

LUCY NWOSU

EDUCATION

PhD. Student in Electrical Engineering, Prairie View A&M University, Prairie View, Texas 77446, USA, 2019 - Present

M.Sc. Computer Engineering, University of Houston, Clearlake, Houston, Texas.

B.Eng. Electronic engineering, University of Nigeria, Nsukka, Nigeria

WORK EXPERIENCE

Graduate Research Assistant, The Center for Computational Systems Biology, Department of Electrical and Computer Engr, Prairie View A&M University, Prairie View, Texas 77446, USA, 2019 - Till date

Technical Consultant, Larson & Toubro Infotech, USA, 2018 - 2019

Graduate Research Assistant, Telecommunication Lab, UHCL, Houston, Texas, 2016 - 2018

PUBLICATIONS

1. Nwosu, Lucy, Xiangfang Li, Lijun Qian, Seungchan Kim, and Xishuang Dong,"Proformer-based Ensemble Learning for Gene Expression Prediction" *in ICIBM 2023 (Accepted)*

2. Nwosu, Lucy, Xiangfang Li, Lijun Qian, Seungchan Kim, and Xishuang Dong,"Calibrated Bagging Deep Learning for Image Semantic Segmentation: A Case Study on COVID-19 Chest X-ray Image", *in PLOS ONE 2022*.

3. Nwosu, Lucy, Xiangfang Li, Lijun Qian, Seungchan Kim, and Xishuang

Dong, "Semi-supervised learning for COVID-19 image classification via ResNet", *in EAI Endorsed Transactions on Bioengineering and Bioinformatics (BEBI), 2021*.

4. Nwosu, Lucy, and Cajetan M. Akujuobi., "Engineering Solutions in the Era of COVID-19 Choosing a CNN Architecture for a Computer-Based Covid-19 Diagnosis", *J Electron and Telecommunication Engineering 1 (2021): 1-11*.

5. Xishuang Dong, Lucy Nwosu, Sheikh Rufsan Reza and Lijun Qian "Effective Screening and Face Mask Detection for COVID Spread Mitigation using Deep Learning and Edge Devices", *Internet of Things - New Insights, 2023 (Submitted, Book Chapter)*

6. Hui Wang, Jiang Lu, Lucy Nwosu, Ishaq Unwala "Two-channel convolutional neural network for facial expression recognition using facial parts", *ssInternational Journal of Big Data IntelligenceVol. 6, No. 3-4 (June 4, 2019), pp 259-268*

7. L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang and T. Zhang, "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts", *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and CST Congress, 2017, pp. 1318-1321*

AWARDS

IEEE Best Poster Presenter Award, 2017 IEEE International Symposium on Dependable Autonomic and Secure Computing (DASC)

IEEE Second Poster Presenter Award, 2017 IEEE 15th Intl Conf on Dependable,Autonomic and Secure Computing, Houston, Texas.