

**APPLICATION OF ARTIFICIAL
INTELLIGENCE IN MICROBIOLOGY AND
CRISPR**

**A THESIS SUBMITTED TO THE INSTITUTE OF
GRADUATE STUDIES OF APPLIED SCIENCES
OF
NEAR EAST UNIVERSITY**

**By
ABDULLAHI IBRAHIM UMAR**

**In Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in
Biomedical Engineering**

NICOSIA, 2021

ABDULLAHI IBRAHIM UMAR

**APPLICATION OF ARTIFICIAL INTELLIGENCE
IN MICROBIOLOGY AND CRISPR**

**NEU
2020**

**APPLICATION OF ARTIFICIAL
INTELLIGENCE IN MICROBIOLOGY AND
CRISPR**

**A THESIS SUBMITTED TO THE GRADUATE
SCHOOL OF APPLIED SCIENCES
OF
NEAR EAST UNIVERSITY**

**By
ABDULLAHI IBRAHIM UMAR**

**In Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in
Biomedical Engineering**

NICOSIA, 2021

**Abdullahi Ibrahim UMAR: APPLICATION OF ARTIFICIAL INTELLIGENCE IN
MICROBIOLOGY AND CRISPR**

**Approval of Director of Graduate School of
Applied Sciences**

Prof.Dr. Hüsnü Can BAŞER

**We certify this thesis is satisfactory for the award of the degree of Doctor of
Philosophy in Biomedical Engineering**

Examining Committee in Charge:

Prof. Dr. Fadi Al-Turjman

Committee Chairman, Research center for AI
and IoT, Faculty of Engineering, NEU.

Fadi Al Turjman

Assoc. Prof. Dr. Enver Ever

Middle East Technical University, Northern
Cyprus Campus

Enver Ever

Prof. Dr. Ahmet Denker

University of South of Florida Department of
Chemical & Biomedical Engineering

Ahmet Denker

Prof. Dr. Mehmet Özsoz

Supervisor, Department of Biomedical
Engineering, NEU

Mehmet Özsoz


Assoc. Prof. Dr. Sertan Serte

Co-Supervisor, Department of Electrical
and Electronic Engineering, NEU

Sertan Serte

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Abdullahi Umar Ibrahim

Signature: 

Date: 28/01/2021

ACKNOWLEDGMENTS

Foremost, I would like to express my warm and heartfelt gratitude to my supervisor Prof. Dr Mehmet Ozsoz for his patience, guidance, motivation, enthusiasm and unwavering support over the course of my doctorate courses and thesis. My deep appreciation goes to my co-supervisor Assoc. Prof. Dr. Sertan Serte for his immense contribution, guidance, support and immense knowledge transfer. I am deeply honored to have worked with Prof Fadi Al-Turjman whose contributions have helped shaped my research and for his unwavering support and guidance on my thesis and published articles. I will also like to appreciate the support and guidance of Dr. Enver Ever, Prof. Dr. Ahmet Denker and Assoc. Prof. Dr. Kamil Dimililer for all their contributions towards the completion of this study.

My warm, heartfelt love and gratitude will always remain with you Prof. Dr. Ayşe Günay Kibarar for all your encouragement, support, motherly love and insightful comments. To my course mates, Pwavodi Conston, Zubaida Amin and Nima Sharma; housemates, Abdulsalam Ibrahim Shema, Salman Muhammad Jeddah, Awwal Muhammad Dawud, Salim Muazu, Ibrahim Mansur, Ahmad Safana friends Polycarp Yakoi and Salahudeen Habib Olakpo, and all of my friends- home and abroad. I will forever remain grateful.

To my farther Prof Umar Ibrahim and my lovely mother Hajiya Shaima'u Abdullahi Bamalli who have shaped me to the man I am today, thank you for believing and supporting me financially, spiritually and morally. My sincere thanks go to my step-mother Hajiya Maryam M Barau for all her support. Finally, to my siblings Usman, Ibrahim, Muhammad, Amir, Ahmadu, Aisha, Rukayya, Maryam and Fatima, thank you for all your prayers and support.

To my parents...

ABSTRACT

Early-stage diagnosis of infectious diseases is crucial for increasing survival rate of patients, preventing spread of disease and decreasing diagnosis and treatment cost. Pneumonia caused as a result of both virus from Coronaviridae family and other viral species such as bacteria, fungi and tuberculosis caused by *Mycobacterium tuberculosis* have been a major challenge for medical and healthcare sectors in many underdeveloped countries and remote communities with limited diagnosis tools and treatment approach. Molecular testing approaches based on anti-body-based approach and PCR techniques are currently the standard techniques employed by medical expert for diagnosis of disease caused by bacteria and viruses.

Other techniques include microscopy, culture test, sputum test, complete blood count. These techniques are still hindered by the utmost need of highly-skilled professionals, equipped and sophisticated tools, chemical reagents and thus, limit point of care diagnosis. Thus, there is need for development of fast, cheap, simple and accurate detection approach for diagnosis and predictions of these diseases. The recent advancement in synthetic biology has led to discovery of Clustered Regular Interspace Short Palindromic Repeat (CRISPR) based on synthetic RNA design to match DNA sequence of pathogens. This technology has proven to be among the most, reliable, accurate, sensitive, specific and fast method for screening of pathogens such as Dengue virus, Zika virus as well as SARS-CoV-2 which is currently under clinical trial.

The fabrication of point of care biosensor based on CRISPR may take time and it will be expensive for underdeveloped countries. Thus, there is high need to explore other alternatives which are simple, fast, reliable, cheap and precise. Pneumonia is among the signs of Covid-19 infection which can be spotted using Chest Xray scan. Tuberculosis can be detected from microscopic slides and chest X-ray but as a result of the high cases of Covid-19 and tuberculosis, this method can be tedious for both Pathologist and Radiologist and can lead to miss diagnosis. This challenge can be solved by employing the means of Computer Aided Detection via AI-driven models which learn features based on convolution and result in an output with high accuracy, precision and recall.

In this paper, we described automated discrimination of X-ray images of disease which are caused by Coronavirus (such as COVID-19) pneumonia caused by different strains of bacteria, non-coronavirus-infection (like as Influenza virus, staphylococcus aureus etc.) and microscopic slide image of tuberculosis into positive and negative cases using pretrained AlexNet Models. Pneumonia images were obtained from optical OCT scans and CXR images with total dataset of 5853. Tuberculosis images were obtained from Near East university hospital. To evaluate the average efficiency of the model, 10K Cross-validation was carried out and machine vs humans. 371 CXR images COVID-19 were obtained from GitHub and Kaggle databases. The model is trained to classify COVID-19 pneumonia from healthy CXR images as well as other binary classification of pneumonia and multiclass (3-way classification and 4-way classification).

For classification of tuberculosis and healthy microscopic slide based on the whole dataset (i.e., 70% for training and 30% for testing), the model achieved training score-accuracy of 99.19%, validation score-accuracy of 98.73%, sensitivity (precision) of 98.59% and specificity (recall) of 98.48%. In terms of cross validation based on 10-K folds, the model achieved average performance of 99.28% training accuracy and 98.29% testing accuracy. To check the efficiency of the model, machine vs human is carried out. 30 unseen images are given to the model, beginners and certified pathologist. The model was able to discriminate the images accurately outperforming both pathologists.

For classification of pneumonia using x-ray images, models are trained based on different splits (50:50, 60:40, 70:30, 80:20 and 90:10) and cross validation based on 10-k folds to differentiate between viral pneumonia and healthy patients. Based on 50:50 data split, the model achieved 97.98% training accuracy, 97.94% testing accuracy, 96.21% sensitivity and 99.00% specificity. 60:40 splits, the model achieved 98.94% training accuracy, 98.95% testing accuracy, 99.09% sensitivity and 98.81% specificity. For 70:30 split, the model achieved 99.19% training accuracy, 98.73% testing accuracy, 98.59% sensitivity and 98.84% specificity. For 80:20 split, the model achieved 99.36% training accuracy, 100% testing accuracy, 99.11% sensitivity and 99.66% specificity. For 90:10 split, the model achieved 99.86% training accuracy, 100% testing accuracy, 99.11% sensitivity and 100% specificity. Based on cross validation, the models achieved average performance of 97.70%

training accuracy, 96.04% testing or validation accuracy-score, 97.34% precision and 97.79% recall.

In terms of discrimination of disease causes by SARS-CoV-2 and other strains of viral species causing pneumonia and normal or negative cases, the trained-model was able to discriminate between different classes based on binary classifications, multiclass (3-way and 4-way classifications) with high accuracies, sensitivities and specificities.

Our result is in line with the notion that CNN models can be used for classifying medical images with higher accuracy and precision. These models can now serve as a confirmation system for diagnosis of both pneumonia and tuberculosis, maximizing miss diagnosis and offer an alternative to relieve the heavy and tedious workload experiencing by radiologist and pathologist in Near East University Hospital. Comparing CRISPR-based biosensors and CAD approaches, CRISPR-based biosensors have shown to be one of the most precise, sensitive and specific approach for diagnosis of infectious disease. CAD on the other hand can be used as a confirmatory approach or as substitute for molecular diagnosis due to it low sensitivity and specificity compare to CRISPR-based method.

Keywords: Pneumonia; Tuberculosis; CRISPR, AlexNet; Cross Validation; X-ray Images; Microscopic slide Images;

ÖZET

Enfeksiyon hastalıklarının erken evrede teşhisi, hastaların sağkalım oranının artırılması, hastalığın yayılmasının önlenmesi, teşhis ve tedavi maliyetlerinin düşürülmesi açısından çok önemlidir. COVID-19'un neden olduğu pnömoni, COVID-19 olmayan virüsler, bakteriler ve Mycobacterium tuberculosis'in neden olduğu mantarlar ve tüberküloz, birçok az gelişmiş ülkede ve sınırlı teşhis araçları ve tedavi yaklaşımına sahip uzak topluluklarda tıp ve sağlık sektörleri için büyük bir zorluk olmuştur. Anti-vücut temelli yaklaşıma ve PCR tekniklerine dayalı moleküler test yaklaşımları, şu anda tıp uzmanları tarafından bakteri ve virüslerin neden olduğu hastalıkların teşhisi için kullanılan standart tekniklerdir.

Diğer teknikler arasında mikroskopi, kültür testi, balgam testi, tam kan sayımı bulunur. Bu teknikler, yüksek vasıflı profesyonellere, donanımlı ve sofistike aletlere, kimyasal reaktiflere olan ihtiyaç nedeniyle hala engellenmektedir ve dolayısıyla bakım noktası teşhisini sınırlamaktadır. Bu nedenle, bu hastalıkların teşhis ve öngörülerini için hızlı, ucuz, basit ve doğru tespit yaklaşımının geliştirilmesine ihtiyaç vardır. Sentetik biyolojideki son gelişmeler, patojenlerin DNA dizisini eşleştirmek için sentetik RNA tasarımına dayanan Clustered Regular Interspace Short Palindromic Repeat'in (CRISPR) keşfedilmesine yol açtı. Bu teknolojinin, Dang virüsü, Zika virüsü gibi patojenlerin yanı sıra halihazırda klinik deneme aşamasında olan SARS-CoV-2 gibi patojenlerin tespiti için en doğru, güvenilir, hassas, spesifik ve hızlı yöntem olduğu kanıtlanmıştır.

CRISPR'ye dayalı bakım noktası biyosensörünün imalatı, gelişmemiş ülkeler için zaman alabilir ve pahalı olabilir. Bu nedenle, basit, hızlı, güvenilir, ucuz ve kesin olan diğer alternatifleri keşfetmeye büyük ihtiyaç vardır. Pnömoni, Covid-19 hastalığının göğüs röntgeni taramasıyla tespit edilebilen semptomlarından biridir. Tüberküloz, mikroskopik slaytlardan ve göğüs röntgeninden tespit edilebilir ancak çok sayıda Covid-19 ve tüberküloz vakası nedeniyle bu yöntem hem Patolog hem de Radyolog için yorucu olabilir ve yanlış tanıya yol açabilir. Bu zorluk, evrişime dayalı özellikleri öğrenen ve yüksek doğruluk, hassasiyet ve geri çağırma ile sonuçlanan yapay zeka destekli modellere dayalı Bilgisayar Destekli Algılama kullanılarak çözülebilir.

Bu yazıda, pnömoninin röntgen görüntülerinin (Bakteriyel, COVID-19 ve COVID-19 olmayan viral pnömoni) ve tüberkülozun mikroskopik slayt görüntüsünün önceden eğitilmiş AlexNet Modelleri kullanılarak pozitif ve negatif vakalara sınıflandırılmasını tanımladık. Pnömoni görüntüleri optik Koherens Tomografi ve göğüs röntgeni görüntülerinden toplam 5853 veri setinden elde edildi. Tüberküloz görüntüleri Yakın Doğu üniversite hastanesinden elde edildi. Modelin ortalama verimliliğini değerlendirmek için 10K Çapraz doğrulama gerçekleştirildi ve makineye karşı insan. 371 CXR görüntüsü COVID-19 GitHub ve Kaggle veritabanlarından elde edildi. Model, COVID-19 pnömonisini sağlıklı CXR görüntülerinden ve diğer ikili pnömoni ve multiclass sınıflandırmasından (3 yollu sınıflandırma ve 4 yollu sınıflandırma) sınıflandırmak üzere eğitilmiştir.

Tüm veri setine göre tüberküloz ve sağlıklı mikroskopik slayt sınıflandırması için (yani eğitim için% 70 ve test için% 30), model% 99,19 eğitim doğruluğu,% 98,73 test doğruluğu,% 98,59 duyarlılık ve% 98,48 özgüllük elde etti . 10-K kıvrımlara dayalı çapraz doğrulama açısından model,% 99,28 eğitim doğruluğu ve% 98,29 test doğruluğu ortalama performansına ulaştı. Modelin verimliliğini kontrol etmek için makine vs insan yapılır. Modele, yeni başlayanlara ve sertifikalı patoloğa 30 görünmeyen görüntü verilir. Model, her iki patologdan daha iyi performans gösteren görüntüleri doğru bir şekilde sınıflandırmayı başardı.

X-ışını görüntüleri kullanılarak pnömoninin sınıflandırılması için, modeller farklı bölünmelere (50:50, 60:40, 70:30, 80:20 ve 90:10) göre eğitilir ve 10-k katlara dayalı çapraz doğrulama, viral pnömoni ve sağlıklı hastalar. 50:50 veri bölüşümüne dayalı olarak, model% 97.98 eğitim doğruluğu,% 97.94 test doğruluğu,% 96.21 hassasiyet ve% 99.00 özgüllük elde etti. 60:40 bölme, model% 98,94 eğitim doğruluğu,% 98,95 test doğruluğu,% 99,09 duyarlılık ve% 98,81 özgüllük elde etti. 70:30 bölme için model% 99,19 eğitim doğruluğu,% 98,73 test doğruluğu,% 98,59 duyarlılık ve% 98,84 özgüllük elde etti. 80:20 bölme için model,% 99,36 eğitim doğruluğu,% 100 test doğruluğu,% 99,11 hassasiyet ve% 99,66 özgüllük elde etti. 90:10 bölme için model,% 99,86 eğitim doğruluğu,% 100 test doğruluğu,% 99,11 duyarlılık ve% 100 özgüllük elde etti. Çapraz doğrulamaya dayalı olarak, modeller% 97,70 eğitim doğruluğu,% 96,04 test doğruluğu,% 97,34 duyarlılık ve% 97,79 özgüllük ortalama performans elde etti.

Covid-19 ve sağlıklı CXR görüntülerinin sınıflandırmaları için model,% 99,71 eğitim doğruluğu,% 99,16 test doğruluğu,% 97,44 hassasiyet ve% 100 özgüllük elde etti. Covid-19 ve Covid-19 dışı viral pnömoninin sınıflandırılması için model,% 99,57 eğitim doğruluğu,% 99,62 test doğruluğu,% 90,63 duyarlılık ve% 99,89 özgüllük elde etti. Covid-19 dışı viral pnömoni ve sağlıklı CXR görüntülerinin sınıflandırılması için model,% 96.43 eğitim doğruluğu,% 94.05 test doğruluğu,% 98.19 duyarlılık ve% 95.78 özgüllük elde etti. Bakteriyel pnömoni ve sağlıklı CXR görüntülerinin sınıflandırılması için model,% 95,28 eğitim doğruluğu,% 91,96 test doğruluğu,% 91,94 duyarlılık ve% 100 özgüllük elde etti. Çok sınıflı 3 yollu sınıflandırma (Covid-19, bakteriyel pnömoni ve CXR görüntüleri)% 97,40 eğitim doğruluğu,% 95,00 test doğruluğu,% 91,30 duyarlılık ve% 84,78 özgüllük ile sonuçlanır. 4 yollu sınıflandırmaya göre (Covid-19 olmayan viral pnömoni, Covid-19, bakteriyel pnömoni ve CXR görüntüleri), model% 94,18 eğitim doğruluğu,% 93,42 test doğruluğu,% 89,18 duyarlılık ve% 98,92 özgüllük elde etti.

Elde ettiğimiz sonuç, CNN modellerinin tıbbi görüntüleri daha yüksek doğruluk ve hassasiyetle sınıflandırmak için kullanılabileceği düşüncesi ile uyumludur. Bu modeller artık hem pnömoni hem de tüberküloz tanısı için bir doğrulama sistemi olarak hizmet edebilir, yanlış tanıyı en üst düzeye çıkarabilir ve Yakın Doğu Üniversitesi Hastanesi'nde radyolog ve patologun yaşadığı ağır ve yorucu iş yükünü hafifletmek için bir alternatif sunar. CRISPR tabanlı biyosensörler ve CAD yaklaşımlarını karşılaştıran CRISPR tabanlı biyosensörler, bulaşıcı hastalığın tespiti için en hassas ve spesifik yöntem olduğunu göstermiştir. Öte yandan CAD, doğrulayıcı bir yaklaşım olarak veya CRISPR tabanlı yönteme kıyasla düşük duyarlılık ve özgüllük nedeniyle moleküler tanı yerine kullanılabilir.

Anahtar Kelimeler: Pnömoni; Tüberküloz; CRISPR; AlexNet; Çapraz Doğrulama; X-ışını Görüntüleri; Mikroskopik slayt Görüntüleri;

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	i
ABSTRACT.....	iii
ÖZET.....	vi
TABLE OF CONTENTS.....	ix
LIST OF TABLES.....	xiii
LIST OF TABLES.....	xv
LIST OF ABBREVIATIONS.....	xi

CHAPTER 1: INTRODUCTION

1.1 Artificial Intelligence.....	2
1.1.1 Machine Learning.....	2
1.1.2 Deep learning and Artificial Neural Networks (ANNs).....	3
1.1.3 Convolutional Neural Networks (CNNs).....	4
1.1.4 AlexNet.....	5
1.1.5 Transfer Learning (TL).....	6
1.1.6 Model parameters: Training, Learning and Testing.....	7
1.1.7 Application of AI in Medicine.....	8
1.2 Thesis Problem.....	9
1.5 Assumption and Limitations of the Study.....	11
1.5 Structure of The Thesis.....	12

CHAPTER 2: CLINICAL BACKGROUND

2.1 Tuberculosis.....	13
2.1.1 Transmission.....	13
2.1.2 Diagnosis and Treatment.....	15
2.2. Pneumonia.....	16
2.2.1 Diagnosis and Treatment of Pneumonia.....	18
2.3 COVID-19 and Pneumonia.....	18

2.3.1 Pandemic and Epidemic of Coronaviridae Family.....	19
2.3.2 Diagnosis and Treatment of Coronavirus.....	21
2.4 CRISPR/Cas System.....	22
2.3.1 Discovery of CRISPR.....	23
2.3.2 CRISPR in Nature.....	26
2.3.3 Classification of CRISPR-Cas system.....	27
2.3.3 CRISPR as Gene Editing Tool.....	31
2.3.4 Application of CRISPR/Cas systems.....	34
2.3.5 CRISPR-based Biosensors.....	35
2.3.6 CRISPR-based Biosensors for Detection of Tuberculosis.....	37
2.3.7 CRISPR-based Biosensors for Detection of COVID-19 Disease.....	38

CHAPTER 3: LITERATURE REVIEW

3.1 Tuberculosis.....	41
2.2 Pneumonia.....	46
3.3 COVID 19.....	48
3.3.2 COVID 19 vs Normal CXR images.....	48
3.3.2 Multiclass (Covid 19, Non-Viral COVID-19).....	50

CHAPTER 4: EXPERIMENTAL SET UP

4.1. Tuberculosis.....	55
4.1.1 Data collection.....	55
4.1.2 Data split.....	58
4.1.3 Data augmentation.....	59
4.1.4 Machine Vs Human.....	59
4.2. Pneumonia.....	59
4.2.2 Data split.....	60
4.3 COVID-19.....	60
4.3.1 Dataset.....	60

4.3.2 Data split.....	62
4.4 Model Training.....	63
4.5 Cross Validation.....	67
4.6 Evaluation and Confusion Matrix.....	67

CHAPTER 5: PERFORMANCE EVALUATION

5.1 Tuberculosis.....	69
5.1.1 Performance Evaluation.....	69
5.1.2 Cross Validation.....	71
5.1.3 Machine Vs Human.....	72
5.1.4 Comparison with state of art.....	72
5.2 Pneumonia.....	73
5.2.1 General Dataset.....	73
5.2.2 Cross Validation.....	75
5.2.3 Comparison between General Data Split and Cross Validation.....	76
5.2.4 Comparison with State of Art.....	77
5.3 COVID-19 and Other Pneumonia.....	78
5.3.1 Binary Classification.....	78
5.3.2 Multiclass.....	79
5.3.3 Comparison with State of Art.....	80
5.3.4 Comparison between CRISPR-based Biosensors and CAD.....	84

CHAPTER 6: CONCLUSION

REFERENCES.....	83
------------------------	-----------

APPENDICES.....	98
------------------------	-----------

Appendix 1: Machine Vs Human 1.....	98
Appendix 2: Machine Vs Human Result.....	100
Appendix 3: Machine Vs Human 2.....	101

Appendix 4: Machine Vs Human Result.....	102
Appendix 5: Codes.....	103
Appendix 6: Ethical Approval Documents.....	105

LIST OF TABLES

Figure 1.1: History of AI, ML and DL.....	2
Figure 1.2: DL architecture with input layer, hidden layers and output layer.....	3
Figure 1.3: Performance of ImageNet architectures.....	4
Figure 1.4: AlexNet architecture.....	6
Figure 1.5: Transfer Learning.....	6
Figure 1.6: Applications of artificial intelligence within the Healthcare Landscape.....	9
Figure 2.1: <i>Mycobacterium tuberculosis</i> (MTB).....	13
Figure 2.2: Pathogens that causes pneumonia.....	15
Figure 2.3: Coronavirus species.....	18
Figure 2.4: Transmission of MERS-CoV.....	19
Figure 2.5: Gene editing Tools.....	22
Figure 2.6: Timeline of CRISPR.....	25
Figure 2.7: Stages involve in CRISPR Adaptive Immunity in Bacteria.....	26
Figure 2.8: Classification of class 2 CRISPR-Cas system based on Domain.....	29
Figure 2.9: The use of CRISPR as Gene Editing Tool.....	32
Figure 2.10: Molecular mechanism of CRISPR/Cas9-mediated DNA cleavage.....	32
Figure 2.11: Cas12a/Cas13a-CRISPR-based Biosensor.....	35
Figure 2.12: HUDSON and SHERLOCK.....	36
Figure 2.13: Detection of <i>Mycobacterium tuberculosis</i> using 2 dCas9.....	37
Figure 2.14: Detection of SARS-CoV-2 using Cas13a/Cas13b.....	38
Figure 4.1: Work flow chart.....	52
Figure 4.2: Positive (tuberculosis) microscopic slide image.....	53
Figure 4.3: Negative Tuberculosis Microscopic Slide.....	57
Figure 4.4: Birme Interfaced.....	58
Figure 4.5: Pediatric CXR scans. Left: Pneumonia. Right: Normal CXR scan.....	60
Figure 4.6: Workflow for Multiclass Classification.....	61

Figure 4.7: Dataset description.....	62
Figure 4.8: Classification of Medical Images using Pretrained AlexNet Model.....	64
Figure 5.1: Classification of TB and healthy MS using pretrained AlexNet Model.....	63
Figure 5.2: Learning Curve of AlexNet network.....	64
Figure 5.3: Performance evaluation of AlexNet models using different data split	68

LIST OF TABLES

Table 2.1: Classification of pneumonia based on Pathogens.....	16
Table 2.2: Discovery and advancement of CRISPR.....	23
Table 2.3: Classification of Cas System.....	27
Table 2.4: Differences between Cas effectors.....	33
Table 3.1: Detection of tuberculosis using AI-driven tools.....	42
Table 3.2: Detection of different types of pneumonia using AI-driven tools.....	45
Table 3.3: Binary Classification of pneumonia using AI-driven tools.....	47
Table 3.4: Multiclass Classification of pneumonia using AI-driven tools.....	50
Table 4.1: Data augmentation of Training Dataset.....	54
Table 4.2: Datasets Description.....	55
Table 4.3: Data split.....	55
Table 4.4: Dataset Description for Binary Classes.....	57
Table 4.4: Dataset Description for Multiclass.....	58
Table 4.6: Confusion matrix.....	61
Table 5.1: Model Learning Parameters and Performance.....	63
Table 5.2: Cross validation result.....	64
Table 5.3: Performance evaluation of machine vs humans.....	65
Table 5.4: Comparison between our result and State-of-the-art.....	66
Table 5.5: Performance evaluation of models trained using different data split.....	67
Table 5.6: Cross validation result for Pneumonia.....	69
Table 5.7: Performance evaluation of models trained using different data split.....	69
Table 5.8: Comparison between our result and state-of-the-Art.....	71
Table 5.9: Performance evaluation of binary classes and multiclass.....	73
Table 5.10: Comparison between our Result (BC) with the State-of-the-Art.....	75
Table 5.11: Multiclass Comparison between our Result with the State-of-the-Art.....	77
Table 5.12: Comparison between CRISPR-based Biosensors and CAD.....	78

LIST OF ABBREVIATIONS

AC:	Accuracy
aM:	Attomolar (10–18M)
ANN	Artificial Neural Network
AI:	Artificial Intelligence
ARDS:	Acute Respiratory Distress Syndrome
ASCas12a:	Acidaminococcus Cas12a
AUC:	Area Under the Curve
BBD:	Big Biomedical Data
BN:	Batch Normalization
BP:	Bacterial Pneumonia
BPr:	Base Pair
CADe:	Computer Aided Detection
CADx:	Computer Aided Diagnosis
CAP:	Community Acquired Pneumonia
CAS:	CRISPR Associated System
CAS-EXARP: reaction	CRISPR/Cas9 triggered isothermal exponential amplification reaction
CPF1 & CPF2:	CRISPR from Prevotella and Francisella 1 & 2
CFR:	Case Fatality Rate
CNN:	Convolutional neural networks
CONV:	Convolution
CRISPR:	Clustered Regularly Interspaced Short Palindromic Repeat
CRRNA:	CRISPR Ribonucleic Acid

CT:	Computer Tomography
CXR:	Chest X-ray
dCas9:	Deactivated Cas9
DL:	Deep Learning
DMD:	Duchene Muscular Dystrophy
DNA:	Deoxyribonucleic Acid
DSDNA:	Double Strand Deoxyribonucleic Acid
DT:	Decision Tree
EAI:	Explainable Artificial Intelligence
FCL:	Fully Connected Layers
fM:	Femtomolar (10–15M)
FM:	Feature Map
FN:	Feature Negative
FRCNN:	Faster-Region based Convolutional Neural Network
FP:	False Positive
FPR:	False Positive Rate
gFET:	Graphene-based Feld Effect Transistor
GMO:	Genetically Modified Organism
GPU:	Graphical Processing Unit
gRNA:	Guide RNA
HEPN:	Higher Eukaryotic and Prokaryotic Nucleotide
HDR:	Homologous Directed Repair
HOLMES:	1-Hour Lowcost Multipurpose Highly Efficient System
HR:	Homologous Recombination
HPV:	Human Papillomavirus

HUDSON:	Heating Unextracted Diagnostic Sample to Obliterate Nuclease
ICU:	Intensive Care Unit
LbCas12a:	Lactinospiraceae bacterium Cas12a
LwCas13a:	Leptotrichia wadei Cas13a
MBTB:	Mycobacterium Tuberculosis
MCTCP:	Montgomery County Tuberculosis Control Program
MERS:	Middle East Respiratory Syndrome
MERS-CoV:	Middle East Respiratory Syndrome Coronavirus
MiR:	Micro RNA
MIT:	Massachusetts Institute of Technology
ML:	Machine Learning
NA:	Nucleic Acid
NASBA:	Nucleic acid sequence-based amplification
NASBACC:	Nucleic acid sequence-based amplification CRISPR cleavage
NGG:	Nucleotide Guanine-Guanine
NHEJ:	Non-Homologous End Joining
NPs:	Nanoparticles
OCT:	Optical Coherence Tomography
PA:	Pretrained AlexNet
PC:	Personal Computer
PCR:	Polymerase Chain Reaction
PFM:	Protospacer Flanking Motif
PAM:	Protospacer Adjacent Motif
PCR:	Polymerase Chain Reaction
PFS:	Protospacer Flanking Site

PM:	Picomolar (10–12M)
PNA:	Peptide Nucleic Acid
PPD:	Purified Protein Derivative
PsmCas13a:	Prevotella sp. Cas 13a
RAM:	Random Access Memory
RF:	Random Forest
RNA:	Ribonucleic Acid
RPA:	Recombinase Polymerase Amplification
RT:	Reverse Transcript
RT-RPA:	Reverse Transcript Recombinase Polymerase Amplification
SARS:	Severe Acute Respiratory Syndrome
SARS-CoV:	Severe Acute Respiratory Syndrome Coronavirus
SF:	Specificity
SFTS:	Severe Fever Thrombocytopenia Syndrome
SgRNA:	Single Guide RNA
SHERLOCK:	Specific, high-sensitivity, enzymatic, reporter, unlocking
SNP:	Single Nucleotide Polymorphism
SRSR:	Short Regularly Spaced Repeat
SSDNA:	Single Stranded DNA
SSRNA:	Single Strand RNA
ST:	Scrub Typhus
SV:	Sensitivity
SVM:	Support Vector Machine
SVR:	Support Vector Regression

TALENS:	Transcription Activator Like Effector Nuclease TracrRNA Transactivating
T:	Thymine
TB:	Tuberculosis
TL:	Transfer Learning
TNs:	True Negatives
TPs:	True Positives
TPR:	True Positive Rate
TracrRNA:	Transactivating RNA
TST:	Tuberculin Skin Test
VP:	Viral Pneumonia
WHO:	World Health Organization
yM:	Yoctomolar (10 ⁻²⁴ M)
ZFN:	Zinc Finger Nuclease
zM:	Zeptomolar (10 ⁻²¹ M)

CHAPTER 1

INTRODUCTION

Computed Aided Diagnosis (CADx) also referred to Computer Aided Detection (CADE) is a computer-based application that assist medical experts in decision-making (Doi, 2007). In healthcare system, medical practitioner uses medical image to evaluate information such as abnormality from images for proper diagnosis. Interpretation of medical images is very critical in medical field due to the fact that any miss-diagnosis can be detrimental (Halalli & Makandar, 2018). Different field of medicine deals with specific type of images such as microscopic slide images by Microbiologist, pathological stain slide by oncologist, CT scans or Chest Xray (CXR) by Radiologist as well as ultrasound and endoscopy (Chen et al., 2013). Nuclear imaging technique gives significant metabolic information as opposed to anatomical information obtained from conventional mammography, ultrasound and magnetic resonance imaging (Histed et al., 2012).

CAD technology revolve around the use of multiple concepts such as medical image processing, computer vision and Artificial Intelligence (AI). The primary function of CAD system is detection of abnormality in medical images such as providing quantified image metrics to compute probabilities of different diagnoses and identification of potential Regions of Interests (ROIs) (Cicerone & Camp, 2019). These techniques have been applied for detection of different grades of tumors (such as colon, prostate, breast and lung cancer) from pathological stain images, detection of *Mycobacterium tuberculosis* from both microscopic slide images and radiographic images, detection of pneumonia from CT scans and diabetic retinopathy. Other application of CAD systems includes diagnosis of Alzheimer's disease, pathological brain detection, coronary artery disease, bone metastases etc. (Halalli & Makandar, 2018).

The first application of CAD technology in medical field is dated back 1990s and still up to this date, the technology has not reached its full potentials. However, the advancement in the field of Deep Learning (DL) has unfold and unwind many twists that hinders its potentials. The use of DL such as Convolutional Neural Networks (CNNs) and Artificial

Neural Networks (ANNs) have revolutionized CAD technology. Currently, there are many AI-driven models that performed better in terms of classification of diseases than medical experts.

Most of the images obtained from medical devices have low quality or low contrast. To enhance their quality, image pre-processing techniques are employed such as filtering, enhancement of image quality through levelling, noise removal, decreasing background artefacts etc.

1.1 Artificial Intelligence

The concepts of AI have been buzzing throughout the last 5 decades as shown in Figure 1.1. Many scholars vary in defining the concept. However, AI is termed as any technique that enables computers to mimic human behavior.

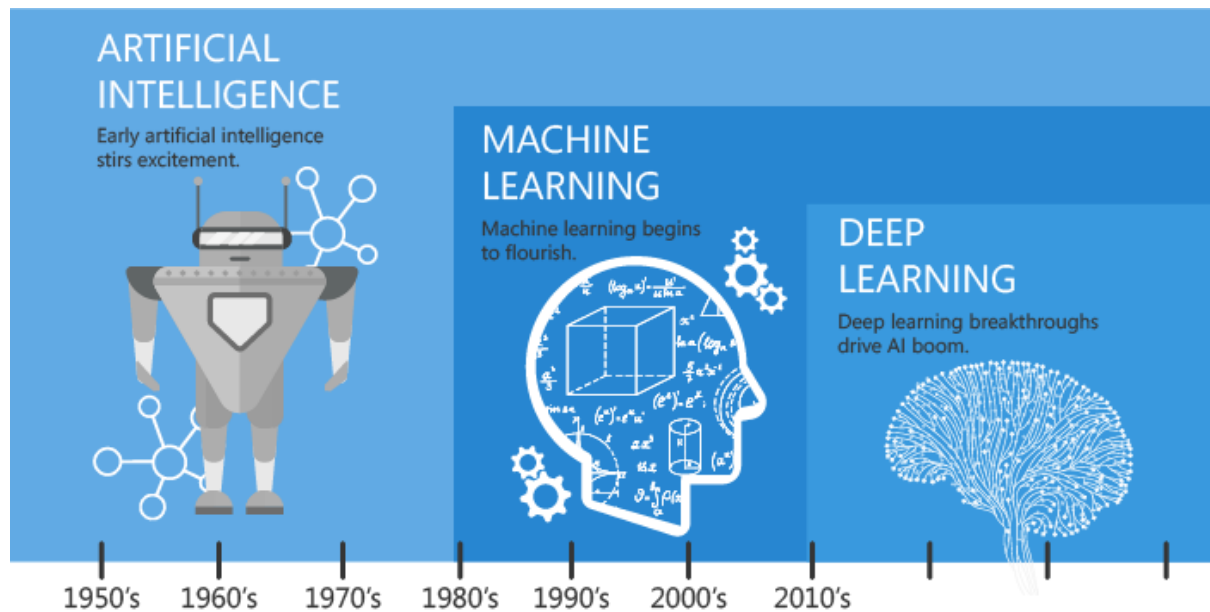


Figure 1.1: History of AI, ML and DL

1.1.1 Machine Learning

ML is defined in 1959 by Arthur Samuel as a field of AI that gives computers the ability to learn without being explicitly programmed. ML algorithms are categorized into 3 main categories, Supervised ML, Unsupervised ML and Reinforcements learning. Supervised ML are the most popular ML techniques utilized by medical expert in which data are labelled and the model or network learn features to identify patterns in data for prediction or

classification. The most widely used supervised ML techniques include Neural Networks (NNs), Support Vector Machine (SVM), Random Forest (RF), decision Tree (DT), etc. (Brownlee, 2016; Paiva et al., 2018).

In Unsupervised ML, models learn from unlabeled data. These models learn to forecast result from the data based on the patterns learned. The most common unsupervised machine learning models include Rule Mining and Clustering algorithms. Subsequently, Reinforcements learning is termed as “when a computer program learns from Experience (E) with respect to some Task (T) based on the Performance (P) measures, if P at T is measured by P improve with experience.

1.1.2 Deep learning and Artificial Neural Networks (ANNs)

Deep learning is a subfield of ML which is inspired by how human brain’s function due to connections or synopsis of nerve cells or neurons. Model learn as a result of data connection between neurons in the network. A simple neural network is termed as perceptron which take input as data set and produced an output as classification category or prediction outcome. Deep learning neural networks are made of multiple perceptron’s with an input layer (IL), and many hidden layers (HL) before output layer (OL) (Abiyev et al 2018; Helwan & Abiyev 2016) as shown in Figure 1.2. Since the emergence of Deep learning in 2010, scientist have designed different models using convolutional neural networks that can classify and analyze medical images such as cancer, tuberculosis, radiological images for diagnosis of diseases (Mnih & Hinton 2009; Helwan et al 2017).

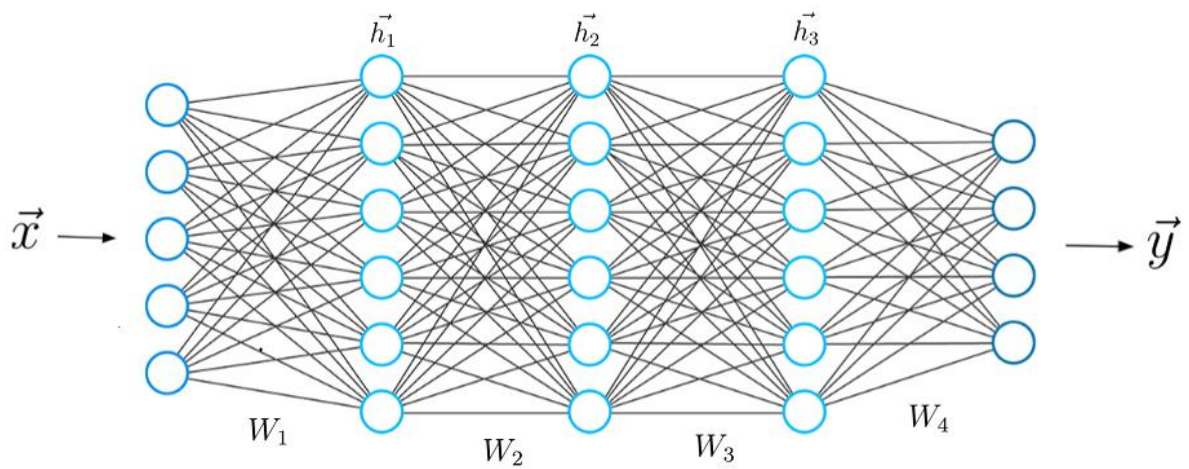


Figure 1.2: DL architecture with input layer (IL), hidden layers (HL) and output layer (OL)

1.1.3 Convolutional Neural Networks (CNNs)

CNN is a class of ANN with multi-layer perceptron which are fully connected network in which each neuron from one layer is connected to all neurons in the next layer. CNN are termed as networks that utilize series of mathematical operations knows as “Convolution” There are various neural networks architectures developed. Some of the architectures have performed better than others in terms of regression, classification and denoising images. The current best models include AlexNet with 8 layers, VGGNet with 19 and 16 layers, Inception module also known as GoogleNet with 22 layers and 9 modules and Residual or ResNet with 152 layers (Russakovsky et al 2015) as shown in Figure 1.3. To train a NN, a backpropagation algorithm is used to adjust the weight according to the data pattern and optimize the error between predicted output and actual output (Simonyan & Zisserman 2014; He et al 2016).

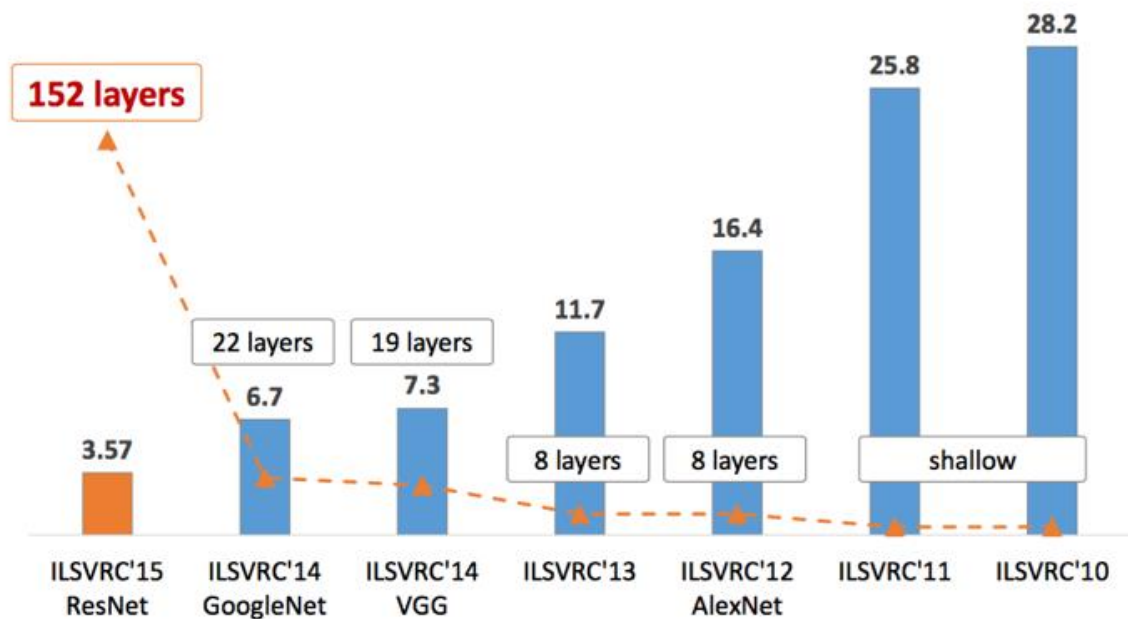


Figure 1.3: Performance of ImageNet architectures

Some of the challenges of these models is underfitting and overfitting. An effective model also known as fitted model is the one that result in high accuracy during training and testing with low bias and variance. Overfitting is a challenge that occurs during training which result

in high training accuracy and low testing accuracy due to high variance and low bias. Underfitting on the other hand is due to low variance and high bias which result in low accuracy in both training and testing operations. These challenges can be avoided or can be overcome through cross validation, early stopping, weights pretraining and penalization, drop out layers etc. (Oyedotun et al 2017; Dawud et al 2019).

The principle behind the application of CNN in classification or regression revolves around series of dot products of weight matrices and input matrix. These processes are categorized into two stages known as Feature learning and Classification. Feature learning is based on the use of convolutional blocks with operations such as convolution which is a process of computing input matrix and feature matrix to obtain a convolve map or feature map, activation operation is the use of activation function (AF) like tanh, sigmoid and Rectified Linear Unit (ReLU) to squash output into zero or within ranges of 0 and 1 or from -1 to 1, pooling operation is carry out to reduce computation by taking the most important part of the convolve map by either max pooling or average (mean) pooling. The output is obtained after these operations in all the layers (including fully connected layers or global average pooling layers) and the use of classifier such as SoftMax based on probabilities to categorized output (Acharya et al 2018; Smith & Topin 2016).

1.1.4 AlexNet

Is the first CNN developed that outperform other models in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in 2012. AlexNet is developed by Alex Krizhevsky. The model consists of overall 8 layers, 5 are convolutional layers and 3 fully connected layers. The first two convolutional layers are made of 3 operations which include convolution, pooling and normalization. AlexNet use ReLU as an activation function unlike Tanh and sigmoid functions that are used in traditional machine learning. ReLU converts negative numbers to zeros and help models learn non-linear functions (Krizhevsky et al 2012; Aloysius et al 2017).

Max pooling is the most common pooling methods which main function is to down sample or to reduce image size by pooling most important feature by pooling out the number with highest pixel value. The next 2 layers are mainly convolution layers without pooling and normalization and the final convolution layer consist of only convolution and pooling

without normalization. The first 2 fully connected layers are dropout layers which main function is to reduce overfitting through reduction of number of neurons. The final FCL is basically for classification as shown in Figure 1.4 (Liang et al 2015; Han et al 2017).

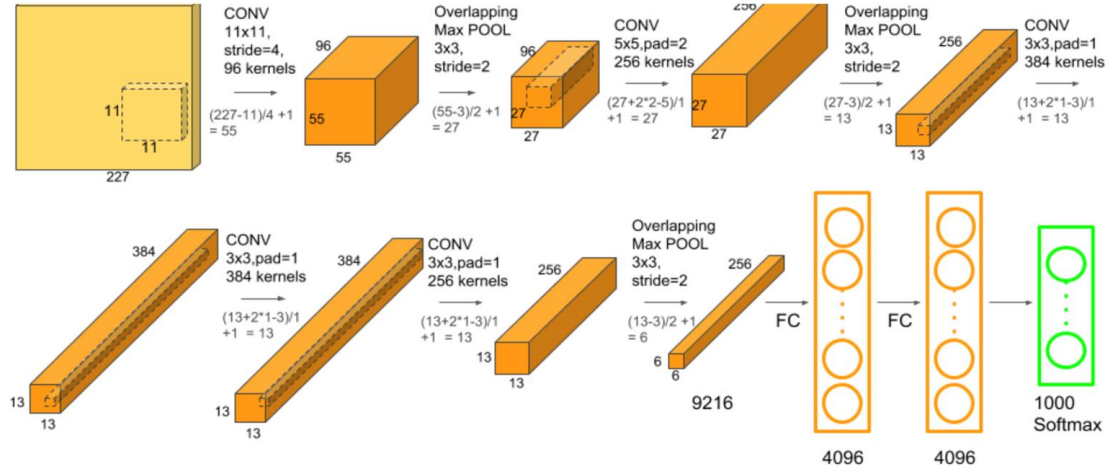


Figure 1.4: AlexNet architecture (Han et al 2017)

1.1.5 Transfer Learning (TL)

TL is defined as ML approach where a model trained on a specific task is re-purposed on other related task or a means to extract knowledge from a source setting and apply it to a different target setting. TL can also be described as a process where what models learned from a specific task or setting is harnessed to improve better outcome in another task or setting as shown in Figure 1.5. TL compare to DL without transfer has demonstrated to be an easier approach that can quickly retrain NNs on dataset of interests with high accuracy (Raghu et al., 2019).

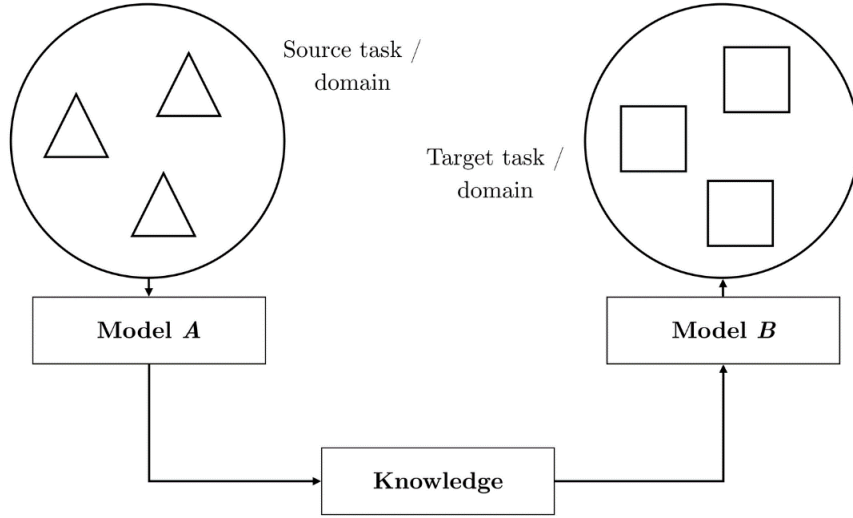


Figure 1.5: Transfer Learning

1.1.6 Model parameters: Training, Learning and Testing

During training, input variables are feed into the network at the input layer which will flow through the hiding layers arriving at the output, as inputs flows between layers they will be multiply by random weights. Different weights signify the intensity of connections between layers and they are the most significant factors in transforming the input to output. Bias and AF is added to the input data as it slides through the model, example of AF is Sigmoid function which will squash any output results between 0 and 1. The obtain outputs will end up been incorrect (i.e., 0.5).

For network to learn, there must be a process of feedback involve which is conducted via back propagation. This process revolves around comparing the output obtained with the output it should produce.

$$O_A - O_O \quad (1.1)$$

The difference will be use to modify the weight of the connection. The network will be run over and over again each time adjusting the weight using Gradient descent which is employed to move the network toward the lowest possible errors. Cost function is a GD which provide a measure of how far off the O_O from O_A . Mean square error is the most common CF use to address the errors.

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (1.2)$$

This equation can be adopted to provide information to the GD that the model needs to reduce the errors across all weight in the network. After training and learning, model can be tested using 25-30% data.

1.1.7 Application of AI in Medicine

Advances in computational technology paired with high amount clinical data and health records has open a window to the use of AI to improve healthcare system. AI has been utilized in healthcare and medicine to increase accuracy of diagnosis, prevent errors, miss-diagnosis, aid in decision making, reducing time of diagnosis etc. Application of AI in healthcare sectors can be categorized into 2 (1) virtual and (2) physical. The virtual application of AI in medicine includes electronic health records system and guidance in decision making using neural networks. In terms of physical application, some of the examples include the use of robots in assisting surgeons in performing surgery and the use of intelligent prostheses for elderly people and handicapped (Amisha et al., 2019; Hamet & Tremblay, 2017; Mintz & Brodie, 2019). In healthcare system, AI have been applied for:

1. **Pharmaceutical:** In pharmacy, AI aid in development of new drugs, identification of potent drugs, selection of patients for specific drug trials, development of vaccines. Scientist have also developed an application known as AI cure that can monitors the use of medications via smartphones etc.
2. **Diagnosis:** diagnosis is the sub-field of medicine that received attention from AI specialist due to the availability of data. Scientist utilize different set of data such as MRI scans (CT and CXR) biopsy tissue samples, microscopic slide images to train computer to distinguish between disease and normal cases (i.e., based on classification or probability score).
3. **Monitoring:** scientist have developed so many AI-driven models that can monitor and track physiological variables such as blood pressure, glucose level, heart rate, sleep levels, intake and suspended calories, temperature etc.
4. **Nursing and AI-Assisted care:** AI specialist and health care professionals developed “Molly” a virtual nurse that can provide follow-up care to patients that are discharged from the clinic so that doctors can focus on in-patients. Different applications are undergoing trials such as AI-Assisted Care designed based on smart ICUs that can

sense any behavioral changes in ICU patients and elderly people living alone. The application of AI is shown in Figure 1.6.

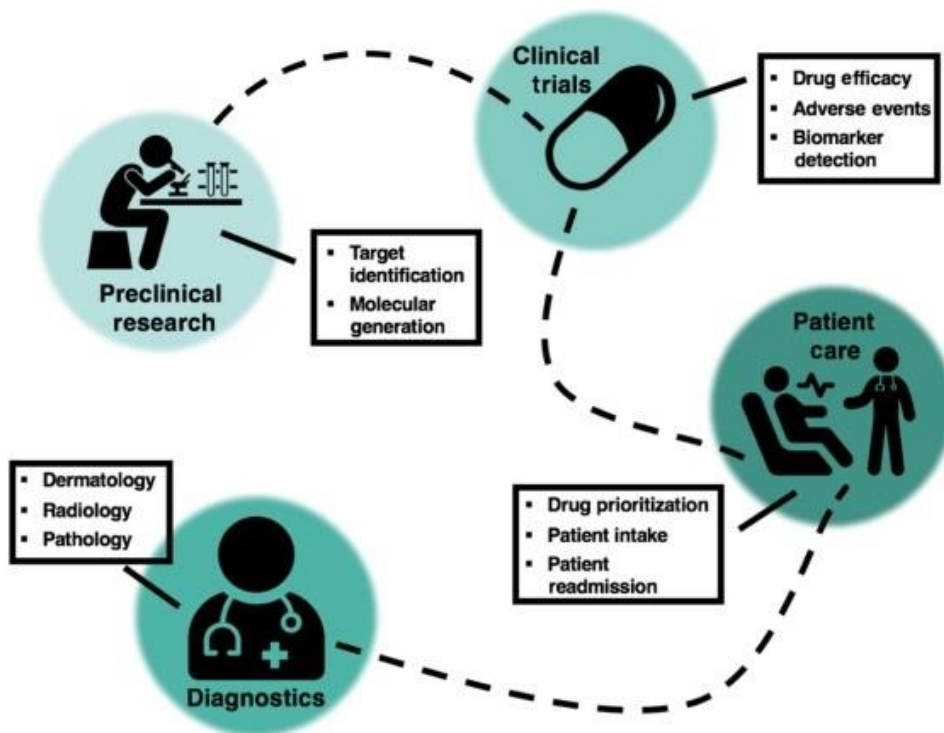


Figure 1.6: Applications of artificial intelligence within the Healthcare Landscape
(Gilvary et al., 2019)

1.2 Thesis Problem

Most of the microbiological, radiological and pathological diagnosis that has to do with stain slides and radiographic images are interpreted by experienced medical experts experienced in the field. However, there have been so many cases of miss-diagnosis, inaccuracy, irreproducibility and false positive results. The use of microscope for quantitative analysis of *Mycobacterium tuberculosis* is very crucial for diagnosis. This approach is clouded by several limitations which can lead to miss-diagnosis (1) The overlapping of the bacteria against each other makes it difficult to carry out accurate diagnosis (2) Examining large amount of microscopic slide image can be very tedious as a result of workload (3) The size of the bacilli is too small as well as its heterogenous shape, irregular appearance, faint boundaries and low background contrast.

Pneumonia have been a major challenge for medical and healthcare sectors in many underdeveloped countries and remote communities with limited diagnosis tools and treatment approach. Interpretation of microscopic slide image and X-ray images of pneumonia obtained from patients is sometimes tedious for qualified pathologist. However, the development of fast, cheap, simple and accurate detection approach for diagnosis and predictions of these diseases are highly required.

As the magnitude for the exponential growth of COVID-19 cases continue to sky rocket so as the number of death tolls around the world. This rise in cases has exhausted many countries testing capacity. Apart from lack of enough test kit, other challenges associated with detection of COVID-19 is the time it takes to transport samples for area of infection to specialized laboratory for RT-PCR testing. This approach (RT-PCR) is the current standard approach used by medical experts for detection of genetic content from pathogens (viruses, bacteria etc.). However, one of the limitations of this method is it provide false positive results which means they are not consistently accurate. In order to solve these challenges, scientists are working beyond the clock to develop a more sensitive, accurate, precise, reliable and point-of-care diagnostic kit that can be utilized for screening or diagnosis of COVID-19 Infection.

The use of CRISPR/Cas systems for detection of pathogens such as Zika virus, dengue virus, pathogenic bacteria such as tuberculosis have been successful in the past. Currently scientists are exploring and harnessing the collateral cleavage of CRISPR/Cas13a in order to develop CRISPR-based biosensor for diagnosis and screening of COVID-19 (aka SARS-CoV-2). Moreover, the fabrication of point of care biosensor based on CRISPR may take time and expensive for underdeveloped countries. Thus, there is high need to explore other alternatives which are simple, fast, reliable, cheap and precise. Considering the fact that pneumonia is one of the symptoms of COVID-19 disease and can be detected using Chest Xray scan, scientist rely on this approach as an alternative or confirmatory approach. However, as a result of high number cases related to COVID-19, this method is tedious to radiologist and can lead to miss diagnosis. This challenge can be solved by using AI-driven models which learn features based on convolution and result in an output with high accuracy, precision and recall.

1.3 Aims and Objectives

- To train DL models to classify microscopic slide image of tuberculosis and normal images.
- To train DL models to classify different types of pneumonia (caused by non-COVID-19 viral pneumonia such as Influenza virus, COVID-19 pneumonia by SARS-CoV-2, bacterial pneumonia from healthy Chest Xray (CXR) images.
- To compare precision in terms of sensitivity for screening of tuberculosis, COVID-19 pneumonia using CRISPR-based biosensors and Deep Learning Models.

1.4 Significance of the Study

The field of medicine and microbiology has witnessed tremendous growth due to invention of microscope which enable scientists to view cells and its organelles. The instruments help medical practitioners to detect pathogenic viruses, bacteria and microscopic fungi. The cases of pneumonia and tuberculosis are predominant in underdeveloped countries with poor healthcare sectors, lack of medical personnel and resources for detection, screening and therapy (Stephen et al 2019; Mathur et al 2018). Thus, these challenges open an area for scientist to developed an alternative, cheap, fast and precise methods that can be used to discriminate between positive cases of specific diseases and normal (negative cases). Thus, the findings of this study will enable pathologist to utilize AI-driven tools as a confirmatory or an alternative approach for the diagnosis of tuberculosis, bacterial pneumonia COVID-19 pneumonia, non-COVID-19 viral pneumonia and normal microscopic slide image or chest Xray scans.

1.5 Assumption and Limitations of the Study

In this study, a pretrained AlexNet model is used for binary classification of tuberculosis from microscopic slide images, COVID-19 pneumonia from normal Chest Xray images and Non-COVID-19 viral pneumonia from normal CT scan images. In terms of pneumonia, different type of pneumonia (bacterial, non-COVID-19 viral and COVID-19) and normal chest Xray images are classified. Apart from AlexNet, there are other high performing CNN models such as ResNet, Inception V3, GoogleNet and VGGNet, but due its simplicity, a

smaller number of layers, minimum error and computational time restraints, it was utilized, nonetheless.

1.5 Structure of The Thesis

The introductory part as Chapter 1 presents the overview of the thesis and concept, including the statement of the problem, the significance of the study, detail aims and objectives and assumptions and limitations. Chapter 2 elaborate on clinical background of tuberculosis, general pneumonia and COVID-19 pneumonia caused by SAR-CoV-2, the chapter also outline symptoms and transmission of the diseases, number of cases, death tolls, diagnosis and treatment as well as CRISPR-based biosensor for detection of tuberculosis and pneumonia. The literature review based on the current-state-of-the arts are addressed in Chapter 3. The methodology of the analysis including dataset and model training are detailed in Chapter 4 with results and discussion addressed in Chapter 5. Conclusion, challenges faced during the analysis and future work are presented in Chapter 6.

CHAPTER 2

CLINICAL BACKGROUND

2.1 Tuberculosis

Tuberculosis is an airborne infection caused by a bacterium known as *Mycobacterium tuberculosis* (MTB) which are slender, rod shape microbes with length ranging from 1-10mm and strict aerobes (need oxygen to survive). They possess a waxy cell wall as a result of formation of “Mycolic Acid” making them “Acid fast” which signifies that they can retain on to a dye or stain in spite of being exposed to alcohol, thus, making the bacilli look red in color when Ziehl-Nelson stain is applied. However, due to the nature of their waxy cell wall, they tend to repelled weak disinfectants and can survive on a dry surface for a longer period of time (Tsai et al., 2013). According to WHO report 2019, deaths as a result of TB-related disease decrease from 1.6 million in 2017 to 1.5 million in 2018. It was estimated that 10 million people fall ill as a result of TB in 2018 with the majority of patients coming from India, Pakistan and china. The symptoms of pneumonia include hemolysis (coughing up blood), fever, weight loss, night sweats etc. (Stewart et al., 2003; Katti, 2004; Polesky et al., 2005).

2.1.1 Transmission

MTB TB can be transmitted via inhalation of droplets from infected person to another through sneezing, coughing, spitting and speaking (Druszczynska et al., 2012). When inhaled, the droplet moves to the lungs, however, the body exercise different defense mechanisms to counter the bacteria:

1. In the upper airways, the air is toxic and thus driving the bacteria against the mucus in order to destroy them. Despite that, the bacteria can maneuverer through the mucus trap and proceed into the lung’s alveoli
2. The lungs contain “Macrophages” which destroy foreign cells. The mechanism behind this activity is as a result of the macrophages recognition of the foreign cells and engulfed the bacteria in a space known as “Phagosome” which contain

lysosomes which are hydrolytic enzyme and can breaks down any biochemical molecules. In spite of this defense mechanism, the bacteria can survive through the engulfment by secreting proteins that can inhibit the action of the enzymes. Thereby causing pulmonary TB which developed symptoms as shown in Figure 2.1. But in some cases, it is asymptomatic with flu-like symptoms.

3. Few weeks after exposure, in response, cells mediated immunity become triggered and immune cells mount on to the site of the infection forming a “Granuloma” in order to prevent the bacteria from proliferating.

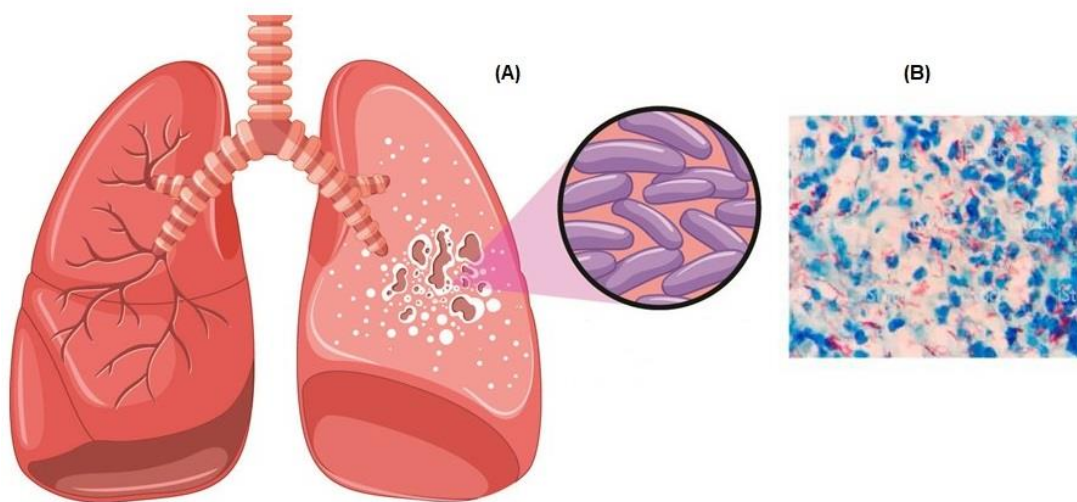


Figure 2.1: (A) *Mycobacterium tuberculosis* (MTB) (B) Ziehl Nelson stain slide showing MTB in Pink color

The tissue inside the granuloma dies due to a process known as “Caseous Necrosis”. The area is termed as “Ghon focus or Ghon complex”. These bacilli can also be transported to close by lymph nodes by the immune cells or through direct extension of the lymph of the Ghon complex leading to caseation. Moreover, the dead-tissue inside the granuloma undergoes a process known as “Fibrosis or calcification” leading to a scar tissue that can be observed using Xray. This calcified Ghon complex is termed as “Ranke complex”. In some cases, these bacilli remain viable or dormant (latent) and can be reactivated when the immune system is weakened due to diseases like HIV and as a result of ageing. The bacilli can spread to either one or both of the upper lobes of the lungs due to the abundance of oxygen. To stop the proliferation of the bacteria. The immune T-cells release cytokines and thus lead to formation of multiple caseous necrosis and form cavities which allows the

bacteria to escape and spread through the lymphatic channels to other regions in the lungs and thus lead to “Bronchopneumonia”. (González-Martín et al., 2010).

When the bacteria spread to other tissues in the body it is termed as “Systematic miliary tuberculosis”. When bacteria spread to other tissues it causes complications to related organs affected such as sterile pyuria of the kidneys due to high levels of white blood cells in the urine, meningitis as a result of meninges of the brain, hepatitis in the liver, Pott diseases associated to the lumbar vertebrae, Addison’s diseases in the adrenal gland, Scrofula as a result of lymphadenitis in the neck etc. (Polesky et al., 2005; González-Martín et al., 2010). The most popular type of TB is pulmonary tuberculosis which attacks the lungs. The other type of tuberculosis is known as extrapulmonary TB which cause damage to other organs such as Kidney, brain and spine. Tuberculosis can lead to mortality if not treated. Another classification of TB is based on active and dormant (latent). Active TB is contagious and can be easily transmitted while latent TB is not contagious but can become active without proper medication (Druszczyńska et al., 2012; Panicker et al 2018).

2.1.2 Diagnosis and Treatment

There are many approaches adopted by pathologist for the detection of tuberculosis, some of the techniques include Tuberculin Skin Test (TST), microscopy, Chest X-ray, Purified Protein Derivative (PPD), GeneXpert, culture test and Interferon γ -release assay. However, among these techniques, microscopic sputum smear evaluation using microscope remain the most common approach globally especially in underdeveloped and developing countries due to its affordability, simplicity, speed and maintenance compare to other techniques (Tsai et al., 2013; Priya & Srinivasan 2015).

The treatment of tuberculosis is highly dependent on the severity of the disease, stage at which it is diagnosed, type of causative pathogens. For latent tuberculosis, doctors prescribe a single drug such as Isoniazid which can be used for a prolonged period of time. For active tuberculosis, treatment depends on the type of causative pathogens, in this case doctors prescribed combinations of different antibiotics. Other measures taken against active tuberculosis include isolation of patients in negative pressure rooms. In the case of drug resistant strains, combination of drugs is prescribed (Katti, 2004; González-Martín et al., 2010; Druszczyńska et al., 2012).

2.2. Pneumonia

Pneumonia is originally derived from Greek word “Pneumon” which means “lungs”. Therefore, pneumonia is attributed to diseases associated with the lungs. In medicine, pneumonia is defined as either inflammation of one or the two lungs parenchyma. Pneumonia occurs due to infections instigated by different types of pathogens which include viruses, bacteria and fungi, exposure to toxic chemicals and food aspiration (Gilani et al 2012). Pathologically, pneumonia infection occurs when the lung’s alveoli became filled up with fluid or puss which in return decrease the exchange of CO₂ and O₂ between blood and the lungs leading to difficulty in breathing. Figure 2.2 and Table 2.1 Shows classification of pneumonia according to pathogens.

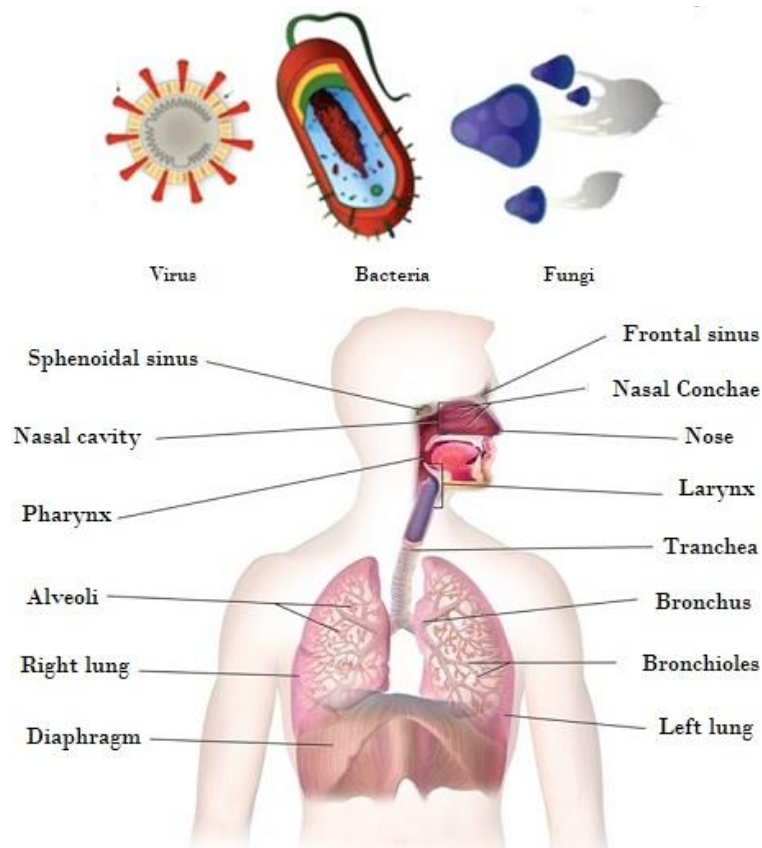


Figure 2.2: Pathogens that causes pneumonia

Table 2.1: Classification of pneumonia based on Pathogens

Pathogen	Specie
Bacteria	<i>S. pneumoniae</i>

	<i>L. pneumophila</i>
	<i>M. pneumoniae</i>
	<i>C. pneumoniae</i>
Viruses	<i>Influenza virus</i>
	<i>Severe Acute Respiratory Syndrome Coronavirus (SAR-CoV-1 and 2)</i>
	<i>Middle East Respiratory Syndrome (MERS) Coronavirus</i>
	<i>Respiratory Syncytial virus (RSV)</i>
	<i>Adenovirus</i>
	<i>Hantavirus</i>
	<i>Rhinovirus</i>
	<i>Varicella-zoster virus</i>
	<i>Human metapneumovirus</i>
	<i>Enteroviruses</i>
Fungi	<i>Pneumocystis jirovecii</i>
	<i>Aspergillus spp</i>
	<i>Mucoromycetes</i>
	<i>Histoplasmosis</i>
	<i>Coccidioidomycosis</i>
	<i>Cryptococcus</i>

The common symptoms of pneumonia include shortness of breath, fever, chest pain, cough etc. Moreover, elderly people (between the age of 65 years and above), children (below the age of 5 years) and patient suffering from other critical diseases such as impaired immune system disorders, patients placed on a ventilator machine, patients with HIV, tuberculosis, asthma, other Acute Respiratory Distress syndromes (ARDS) and other chronic diseases, as well as people who smoke are regarded as the categories prone or at more risk of contracting pneumonia (Falade et al 2011).

According to WHO, over 4 million premature deaths occur due to infections associated with household air pollution including pneumonia and tuberculosis. More than 150 million people were estimated to be infected with pneumonia annually and more prevalence in children less than 5 years old (WHO, 2018). Globally, pneumonia is among the top disease that affect children and account for 15% of mortality of infants and children below 5 years leading to

over 1.4 million death in 2018 and 2.56 million in 2017. Even though the prevalence of the disease is common in children but it can also affect young and older people.

2.2.1 Diagnosis and Treatment of Pneumonia

Medical expert such as clinical pathologist, microbiologist and radiologist employ different method for the diagnosis of pneumonia. Some of these methods includes: sputum test, complete blood count, chest X-ray, Thoracentesis, arterial blood gas etc. the treatments of pneumonia are mostly dependants on the causative pathogens. For viral pneumonia (ones caused by influenza virus, hantavirus, SAR-CoV-1 and 2, MERS etc.) antiviral drugs are used which inhibit or destroy the virus. In the case of bacterial pneumonia (ones caused by *Streptococcus pneumoniae*, *Legionella pneumophila* etc.) antibacterial drugs such as antibiotics are used to destroy or inhibit the growth and differentiation of the bacterial cells. For fungal pneumonia (ones caused by *Pneumocystis jirovecii*, *Aspergillus spp*, *Mucoromycetes* etc.) antifungal drugs are used.

2.3 COVID-19 and Pneumonia

COVID-19 has been making headlines throughout 2020 as a result of global health crisis due to the outbreak of SARS-CoV-2 on the eve of 2020 in Wuhan China. Prior to the year of outbreak, other viruses from same family “Coronaviridae” name SARS-CoV-1 and MERS-CoV have led to epidemics with higher mortality rates (Huang et al., 2020; Gomez et al., 2020). The word “Corona” means “Crown” and thus the virus is named coronavirus after the spike proteins present on their surface which looks like a crown as shown in Figure 2.3.

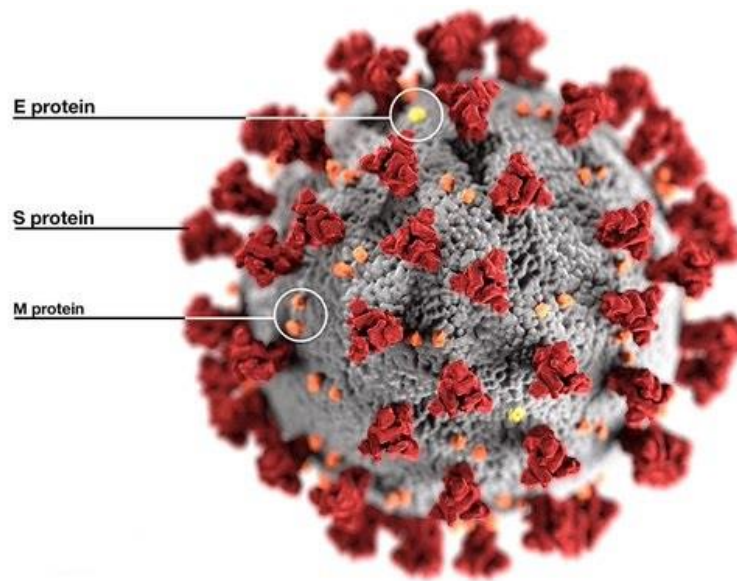


Figure 2.3: Coronavirus specie (Fan et al., 2019)

Like other ARDS, COVID-19 disease can be transmitted in the form of respiratory droplets (sneeze or cough) exhaled by infected person or through physical contact with infected person or coming in contact with surface contaminated by the virus. Just like the pathology of common pneumonia, SARS-CoV-2 when contacted invaded the alveoli present in the which are organs that are liable for the exchange of O₂ and CO₂ between lungs and blood and thus leading to COVID-19 pneumonia. Symptoms associated to COVID-19 disease include: severe pneumonia, dry cough, fever, sputum secretion, fatigue, anorexia, dyspnea, organ failure, myalgias, septic shock etc (Fan et al., 2019; Gomez et al., 2020).

2.3.1 Pandemic and Epidemic of Coronaviridae Family

1. SARS-COV-2 Pandemic

The word “Pandemic” was introduced in 2005 and it was only declared global health emergency on 5 counts, in 2009 as a result of Influenza virus, 2014 as a result of Ebola and polio virus, in 2016 as a result of Zika virus, 2019 as a result of Ebola and in 2020 as a result of SARS-CoV-2. The first case of this virus was declared in a city known as Wuhan, Hubei province in the mainland China on the eve of 1st January, 2020. Though, it was not until 11th March 2020 that the WHO declared the disease “Pandemic”. The virus has spread to over 200 countries and infected more than 30 million and closed to 1 million death toll and still counting.

The mortality rate of SARS-CoV-2 is low compared to SARS-CoV-1 and MERS-CoV, currently, close to 20 million patients have recovered. The outbreak of COVID-19 has led to city lockdowns in many countries, border closure, flight cancellations, restrictions, quarantine, evacuation, cancellations and postponement of educative, sportive, cultural and religious activities. To control the widespread of the virus, government implement the use of facemask, hand sanitation and social distancing based on personal hygiene (Huang et al., 2020).

2. MERS-CoV Epidemic

The first case of MERS-CoV was declared in the month of April 2012 in Saudi Arabia. The virus has spread to 27 countries with majority in the middle east and 8 countries from Europe. According to WHO, prior to the year 2020, there are 2500 laboratory confirmed cases of MERS with close to 1000 death leading to 34% mortality rate. Even though the exact source of the virus is still unknown but scientist attributed the virus to dromedary camel as the main reservoir which transmitted the virus to humans as shown in Figure 2.4. The main symptoms of MERS include shortness of breath, fever, cough, diarrhea etc. (WHO, 2019; CDC, 2019).

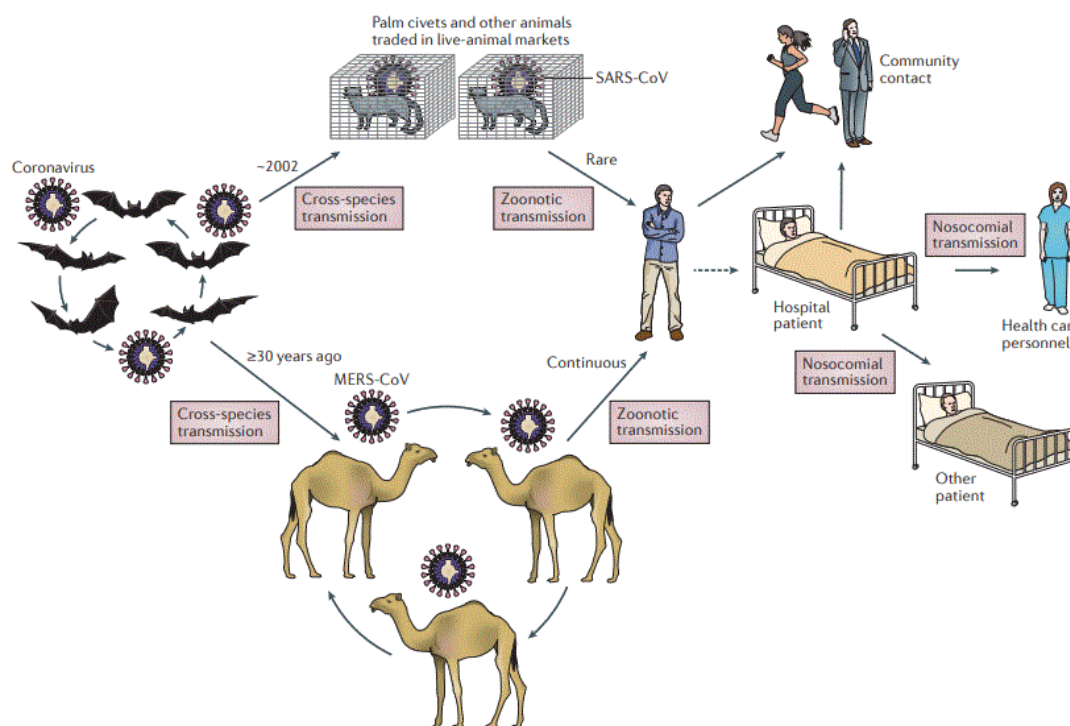


Figure 2.4: Transmission of MERS-CoV

3. SARS-COV-1 Epidemic

The first case of coronavirus (SARS-CoV-1) was declared on the month of November 16, 2002 in Guangdong province of Southern China. Unlike MERS-CoV, bats are regarded as the reservoirs of the virus and civets' cats are the main carriers that infect humans. The symptoms of SARS include diarrhea, fever, dry cough, myalgia, malaise, headache and rigorous shivering among others. However, not all patients suffering from the disease are symptomatic in the first week of illness especially patients who are on immunosuppressants. The virus has spread to 29 countries with over 8 thousand confirmed cases and 774 death tolls worldwide and estimated mortality rate of 9.675% CFR ((Keogh-Brown and Smith, 2008).

The mortality rate of the 3 coronaviruses is shown in Figure 2.5. Mortality rate of infection are calculated based on Infection Fatality Rate (IFR) which is define as the number of deaths over number of infected patients multiply by 100 and Case Fatality Rate (CFR) define as the number of deaths over number of confirmed cases multiply by 100 (WHO, 2019).

2.3.2 Diagnosis and Treatment of Coronavirus

Reverse Transcript Polymerase Chain Reaction (RT-PCR) is a molecular genetic test which is the current gold standard method for detection or diagnosis of viral diseases such as HIV/AIDS, COVID-19, diseases caused by Influenza virus etc. This method directly detects the pathogenic virus through series of steps and procedures. These steps include

Step1: Sample collection based on using swab to obtain samples from suspected person's nose or throat.

Step 2: Extraction: different chemical reagents are added to the sample in order to remove biochemical substances such as fats and proteins to obtained pure viral RNA.

Step 3: Reverse Transcripts (RT): since SARS-CoV-2 is a virus which stored it genetic materials in RNA, scientist converts the RNA to DNA so that it can be amplified.

Step 4: PCR amplification and detection: DNA resulted from RT is amplified using RT-PCR machine to obtain thousands of copies for easy detection (due to the large quantity of the DNA). Detection of the virus is based on the fragments that attach to the target part of

the viral DNA with marker labels which release fluorescent dye which can be quantified by the machine and display the result on the screen (Tahamtan & Ardebili, 2020).

Apart from RT-PCR approach, scientist also use Computed Tomography (CT) scan or Chest Xray scans for screening or diagnosis of COVID-19 pneumonia. The use of CXR and CT scans are the 2 main techniques employed by Radiologist to differentiate between patient suffering from or positive of pneumonia disease and the healthy ones. The discrimination is due to the presence of white hazy patches described as “Ground-glass opacity” in individual positive of pneumonia which is not present in healthy person (Shan et al., 2020; Yang and Yan, 2020).

Currently, there is no specific or agreed drugs or vaccines for the treatment of COVID-19 disease. However, medical experts rely on antiviral drugs that prove effective against other viral diseases. The use of some of these drugs are based on identification AI-driven models which screen thousands of drugs and predict the most potent ones and their probability for use against the virus. Scientist all over the world are working at an unprecedented speed for development of safe and effective vaccines that can prevent the transmission chain of the virus. Currently, there are over 150 candidate vaccines under development with over 20 going through clinical human trials (WHO COVID-19 Vaccines, 2020)

2.4 CRISPR/Cas System

Application of genetic engineering is crucial for correcting genetic disorders such as sickle cell anemia, cystic fibrosis, Duchene muscular dystrophy, Huntington’s disease etc. In the past, scientist have relied on genetic tools such as Zinc Finger Nuclease (ZFN), TALENS, recombinant DNA, meganucleases adenovirus as shown in Figure 2.5. However, these tools are clouded by several challenges such as high off-targets rate, tedious laboratory work, longer period of time, less sensitive, inaccurate and expensive. The discovery of CRISPR in 1987 by Japanese scientist has given birth to a new approach. However, the breakthrough of CRISPR as genetic tool come as a result of experiment carried out by Jennifer Doudna and Emanuel Charpentier in 2012. Unlike previous gene editing tools, CRISPR is cheaper, faster, efficient and easy to use.

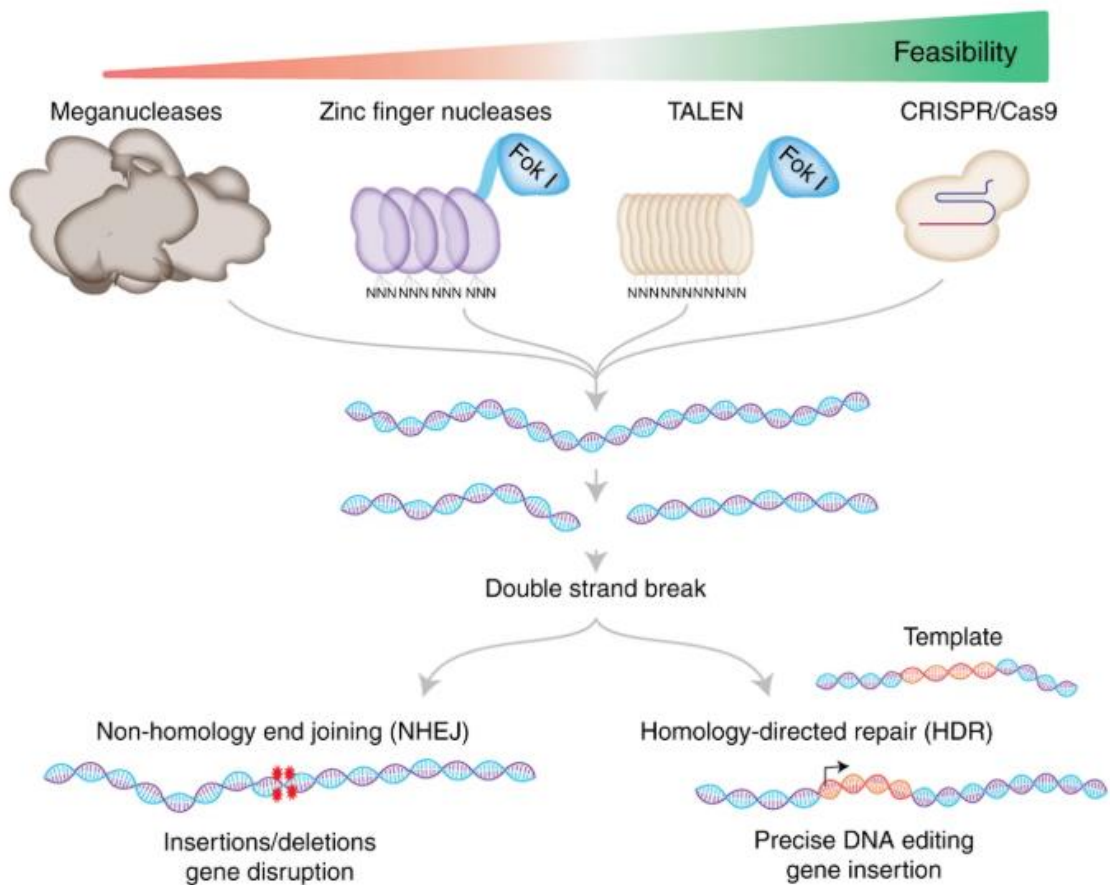


Figure 2.5: Gene editing Tools

CRISPR is an acronym which stand for Clustered Regularly Interspaced Short Palindromic Repeat which is a bioweapon ceased from viruses such as bacteriophage and store in a warehouse known as CRISPR array. These bioweapons are used by bacterial cells as form of immunity to fight foreign invaders. Scientist biomimic this system to developed gene editing tool for editing genes and bases in eukaryotic cells.

2.3.1 Discovery of CRISPR

CRISPR discovery is first dated to the late 1980s when Japanese scientist known as Ishino Yoshizumi, was experimenting using *E. coli* for specific gene responsible for conversion of isoenzyme to alkaline. In the process he discovered an odd sequence which is palindromic (read in both forward and backward direction), he published the work without discussing about the function of the sequence. In 2002, Jansen rename the palindromic sequence as SRSR (Short Regulatory Spaced Repeat). Another group of scientists discovered Cas genes

known as CRISPR associated systems. These genes form a complex structure with guide RNA and together locate target sequence and destroy it (Baltimore et al., 2015).

Between the year 2005-2007, similar sequence was discovered in archaea by Francisco Mojica. The remarkable discovery of CRISPR as an adaptive immune system utilized by bacteria and archaea was achieved by Barrangou and Horvath in 2007 as a result of an experiment using bacteria used in production of yoghurt (Barrangou et al., 2007; Liu et al., 2017). However, the golden key to the Pandora box of CRISPR as genetic engineering tool was discovered Jennifer Doudna and Emanuel Charpentier who together engineered the first artificial single guide RNA (sgRNA) which is a duplex (chimeric) RNA made up of CRISPR RNA (crRNA) and Trans-activating RNA (tracrRNA).

As a result of their discovery, scientists have reported the application of CRISPR/Cas systems to edit prokaryotic cells such as bacteria and archaea. The application of this technology to edit mammalian cells was first carried out by Feng Zeng in 2013 (Barrangou et al., 2007; Ishino et al., 2018). Scientists that contribute to the discovery and advancement of CRISPR are presented in Figure 2.6. and Table 2.2.

Table 2.2: Discovery and advancement of CRISPR

Year	SCIENTIST	DISCOVERY
1987	Yoshizumi Ishino	Discovered CRISPR in <i>E. coli</i>
2000	Mojica and colleagues	CRISPR is present throughout prokaryotes and archaea
2002	Jansen et al	Proposed the word CRISPR
2005	Mojica and colleagues	Sequence between repeat is surprisingly foreign
2007	Alexander Bolotin	Discovery of Cas9 and PAM
2007	Phillipe Horvath	Bacteria are immune to virus and stored a Viral DNA into their CRISPR
2010	Sylvain Morneau	Cas9 is guided by spacer Cas9 cleaves target DNA

2011	Emmanuelle Charpentier	Discovery of tracrRNA for Cas9 system
2012	Charpentier and Doudna	crRNA and the tracrRNA could be fused together to create a single synthetic guide for gene editing
2013		The Use CRISPR-Cas9 for genome editing in eukaryotic cells
2014	Nishimasu and Feng Zhang	Crystal structure of Cas9 gRNA complex

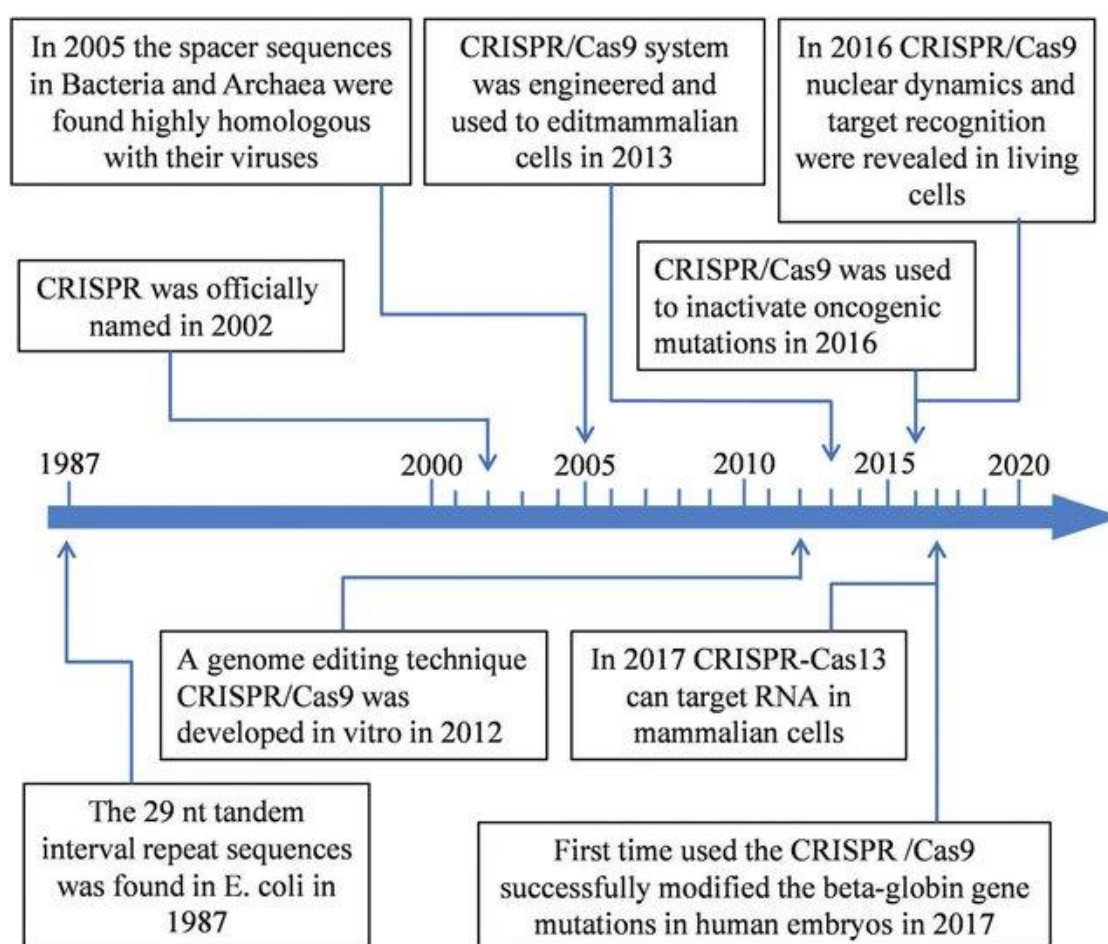


Figure 2.6: Timeline of CRISPR (Wang et al., 2018)

2.3.2 CRISPR in Nature

CRISPR/Cas system is an adaptive immune system that result through evolution in bacteria and archaea as defense mechanism against viruses (Horvath and Barrangou, 2010; Liu et al., 2017). When viruses invade bacteria, they inject their DNA which replicate and destroy the bacterial cell, in order to prevent the virus from replicating, these organisms utilize 3 steps process to ensure immunity (Bhaya et al., 2011; Ishino et al., 2018). These steps or stages include adaptation, expression (biogenesis or recognition) and interference as shown in Figure 2.7.

Adaptation stage occur when bacterial or archaeal cells first came in contact with viral DNA. The CRISPR loci translated Cas genes into Cas proteins (Cas9, Cas2 and Cas1). These Cas proteins surveys for the viral DNA, cut part of it (known as spacer) and store it in CRISPR array's leader strand. The second stage known as expression stage is only initiated when the viral DNA attack again. The bacterial or archaeal cell's CRISPR array transcribed it stores spacers into small non-coding RNA known as Pre- CRISPR RNA which link with TracrRNA through base pairing and form hybrid RNA or matured CRISPR RNA. The CRISPR RNA is employed in the third stage known as interference stage where it forms complex with effector Cas enzyme (such as Cas9 which is translated from the Cas genes adjacent in CRISPR loci). This complex locates the viral DNA as a result of the unique Protospacer Adjacent Motif (PAM) sequence and destroy the viral DNA leading to complete immunity (Barrangou et al., 2007; Garneau et al., 2010; Bhaya et al., 2011).

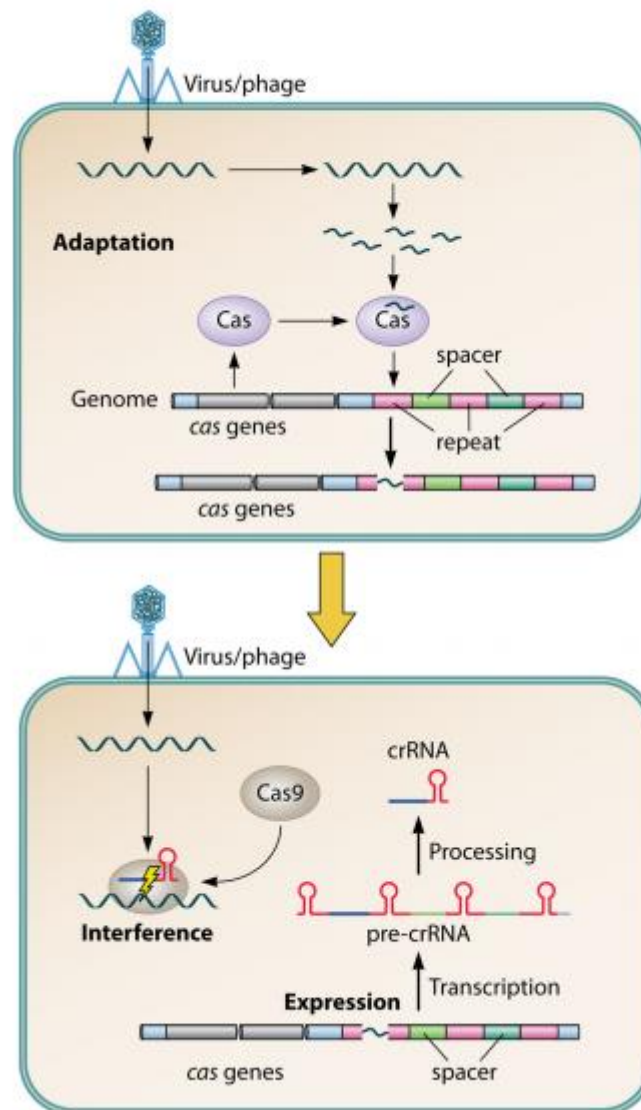


Figure 2.7: Stages involve in CRISPR Adaptive Immunity in Bacteria

2.3.3 Classification of CRISPR-Cas system

The classification of CRISPR-Cas system still remains a challenge to scientist as many types and subtypes are yet to be identified. Throughout the last decade different types and subtypes are identified. CRISPR-Cas systems are classified basically based on adaptor module (i.e., acquisition) and effector module (i.e., maturation of CrRNA, recognition and interference) (Rath et al 2015). Other approaches classify CRISPR-Cas system into 3 main types; type I, type II and type III quite number of subtypes based on genetic content, function and structure. These types differ based on their biochemical functions and the protein they

encode. The universal Cas System which includes Cas 1 and Cas 2 are found in all types and subtypes (Barrangou et al 2014) as shown in Table 2.3.

The recent classification of Cas system is based on the configuration of effector modules. These approaches classify CRISPR-Cas system into class 1 and class 2. Class 1 system combine with CrRNA and other Cas proteins to form effector complex. While class 2 system utilizes a CrRNA along with a large single Cas protein and interfere with target sequence (Zotsche et al 2015). Bioinformatics characterization study of type II CRISPR-Cas system served as the base for discovery of other types, subtypes and classes (Koonin et al 2017).

Table 2.3: Classification of Cas System

Class	Type	Adaptation	Pre-CrRNA Processing	Effector module	Target cleavage
Class 1	I	Cas1, Cas2 and Cas4	Cas 6	Cas7 and Cas5	Cas 3
	III	Cas 1 and Cas 2	Cas 6	Cas7 and Cas5	Cas 10
	IV	-	-	Cas7 and Cas5	-
Class 2	II	Cas1, Cas2 and Cas4	RNaseIII	Cas9	Cas9
	V	Cas1, Cas2 and Cas4	-	Cpf1 (Cas 12)	Cpf1 (Cas 12)
	VI			Cas 13	Cas 13

The class 1 CRISPR-Cas system is made up of type I, type III and type IV and which are mostly found in Archaeal CRISPR loci and less frequent in bacterial cell (Koonin et al 2017). Type I contain different subtypes ranging from type IA-IF. The common Cas candidate of type I is Cas 3 protein which is made of helicase and DNASE domains required for cleaving a specific target. Cas 3 along with other universal Cas system (Cas1 and Cas 2) encode for a complex known as Cascade which bind to CrRNA, unwind the DNA in order to locate the target specific sequence (Devashish et al 2015). Type III is the only Cas system that can target both DNA and RNA. Cas 10 is type III candidate which function is not yet clear (Rath et al 2015).

The class 2 CRISPR-Cas system is made up of type II, type V and type VI as shown in Table 2.4. Type II I known to contain two different subtypes, type II-A and type II-B. Type IICRISPR Cas system is the most adopted Cas system use in genetic engineering (gene

editing). It encodes the universal Cas system (Cas 1 and 2) Cas9 and in some cases Cas 4 (csn2). Cas 9 gene is transcribed to Cas 9 enzyme which along with universal Cas function in adaptation stage. It forms a complex structure with CrRNA and trans-activating RNA (TracrRNA) to cleave target DNA and destroy it (Rath et al 2015).

Cas 9 is the sole effector protein of type II Cas system. The structure and length of Cas 9 varies among different species. Cas 9 is large with several nuclease domains with size ranging from 950 to 1600 amino acids. The two significance domains in Cas 9 are the RuvC domain and HNH domain which are specific for inducing double strand break on target. Cas 9 is the most widely engineered nuclease employed by scientist to edit genome in both prokaryotic and eukaryotic cells (Shmakov et al 2015).

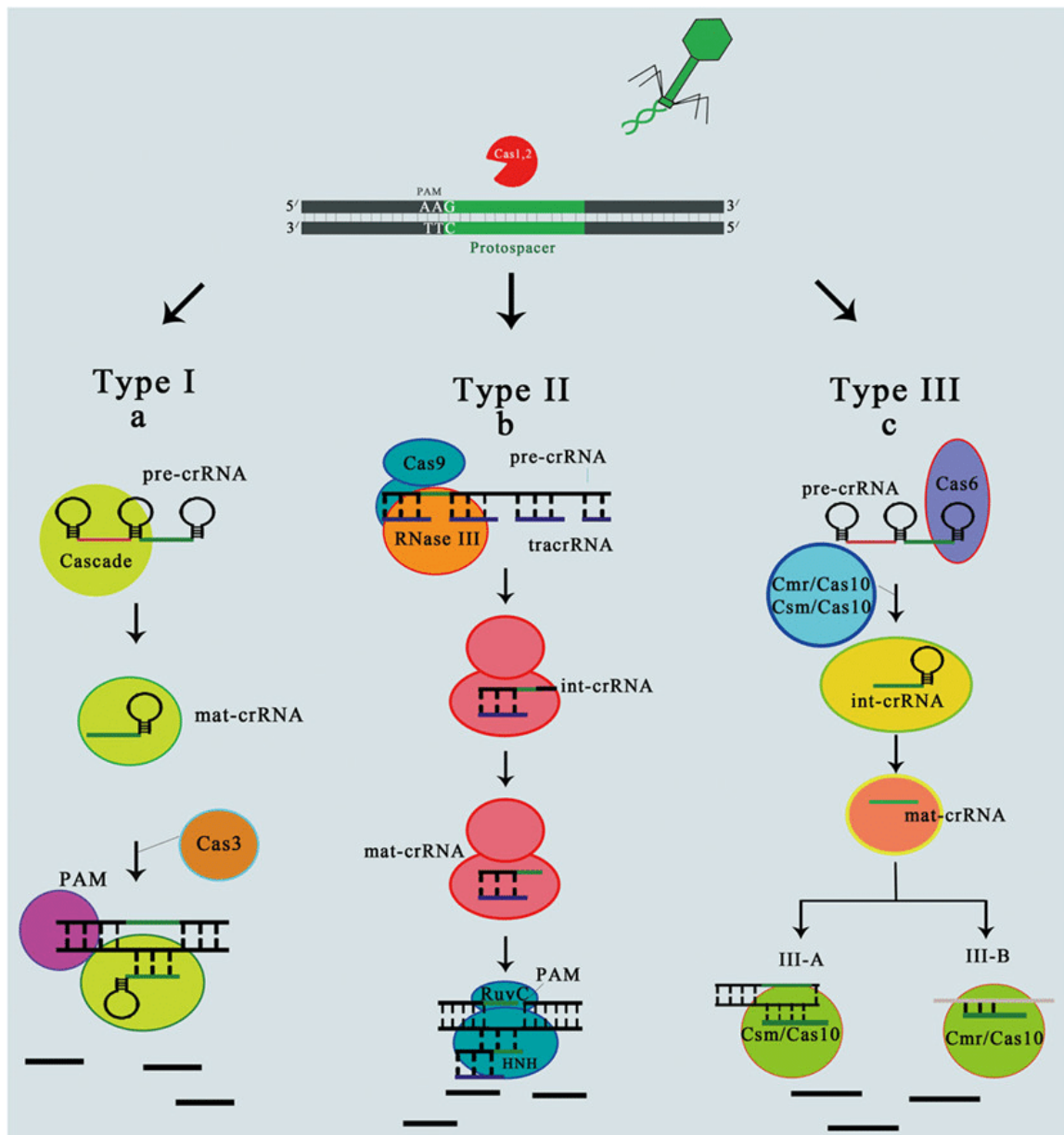


Figure 2.8: Classification of class 2 CRISPR-Cas system based on Domain (Koonin et al 2017)

Cpf1 (Cas 12) is a unique type II Cas system that induced a single break on target due to its single RuvC nuclease domain. It is a single stranded target nuclease that induced a staggered cut on target DNA (Shmakov et al 2015). Cpf1 is an acronym for CRISPR obtained from *Prevotella* and *Francisella*. It is made up of approximately 1,300 amino acids. The uniqueness of Cpf1 is unlike other Cas systems; it does not require trans-activating (TracrRNA). Unlike Cas 9, which guides the target sequence based on NGG PAM, Cpf1 recognizes a T-rich PAM and

cleaved the target or induced a staggered break on the target (Zotsche et al 2015). Type II-A Cas system poses an extra Cas gene known as *csn2* which is mainly employed in adaptation stage but not needed in interference stage. Instead of *csn2* type II-B poses Cas gene known as Cas 4. In all the subtypes of type II, they both poses pair of Cas 1 and Cas 2 genes that work primarily in the acquisition stage (Chylinski et al 2014).

2.3.3 CRISPR as Gene Editing Tool

CRISPR/Cas systems as the new born gene editing tool has revolutionized genetic engineering due to its ability to precisely edit genes. This system is made up of Cas proteins (Cas9, Cas12 and Cas13) which cut DNA/RNA (i.e., Cas9 which acts like a pair of molecular scissors) and synthetic guide RNA (which is made up of 100 nucleotides with the first 20 nucleotide function in navigating the complex to the target sequence) which can be program to target specific sequence and induced double or single strand break. Once the DNA/RNA is cleaved, the cells initiate repairing mechanism known as Non-homologous Enjoining (NHEJ) which is a natural way cells stick together through insertions or deletions of nucleotides (known as indels). However, this method is prone to mutation and can lead to gene dysfunction or deactivation. Scientist can use this window to introduced a desired homologous DNA template through a process known as Homologous Repair (HR) or Homologous Directed Repair (HDR) (Doudna and Charpentier, 2014; Ma et al., 2015) as shown in Figure 2.9.

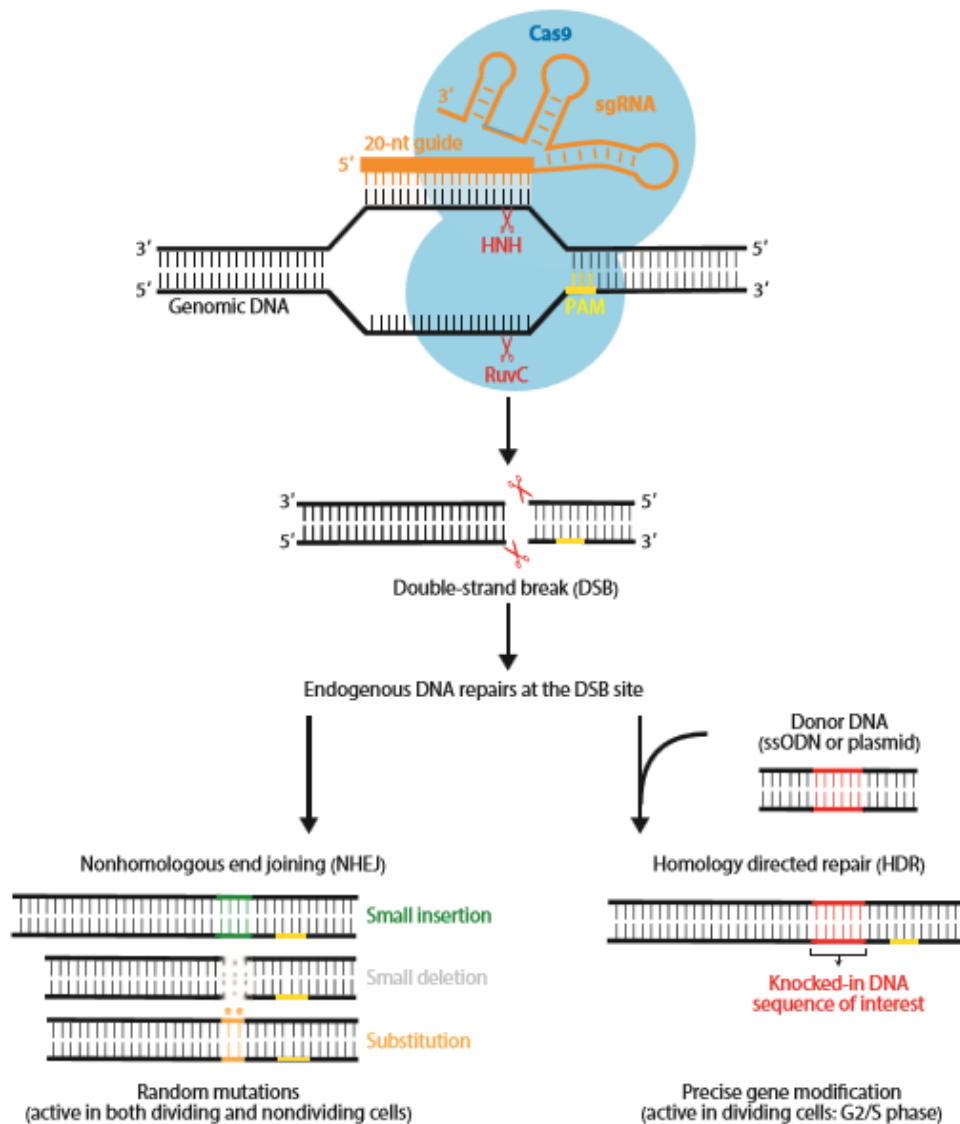


Figure 2.9: The use of CRISPR as Gene Editing Tool

A. CRISPR/Cas 9

CRISPR/Cas9 is the most popular gene editing tool use by scientist to make precise changes in genes of prokaryotic and eukaryotic cells. Cas9 can be isolated from different bacterial species such as *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Staphylococcus aureus* etc. Cas9 enzyme is made up of 2 domains, RubC and HNH. These domains cleaved both opposite and complementary strand of the target sequence (Ma et al., 2015) as shown in Figure 2.10.

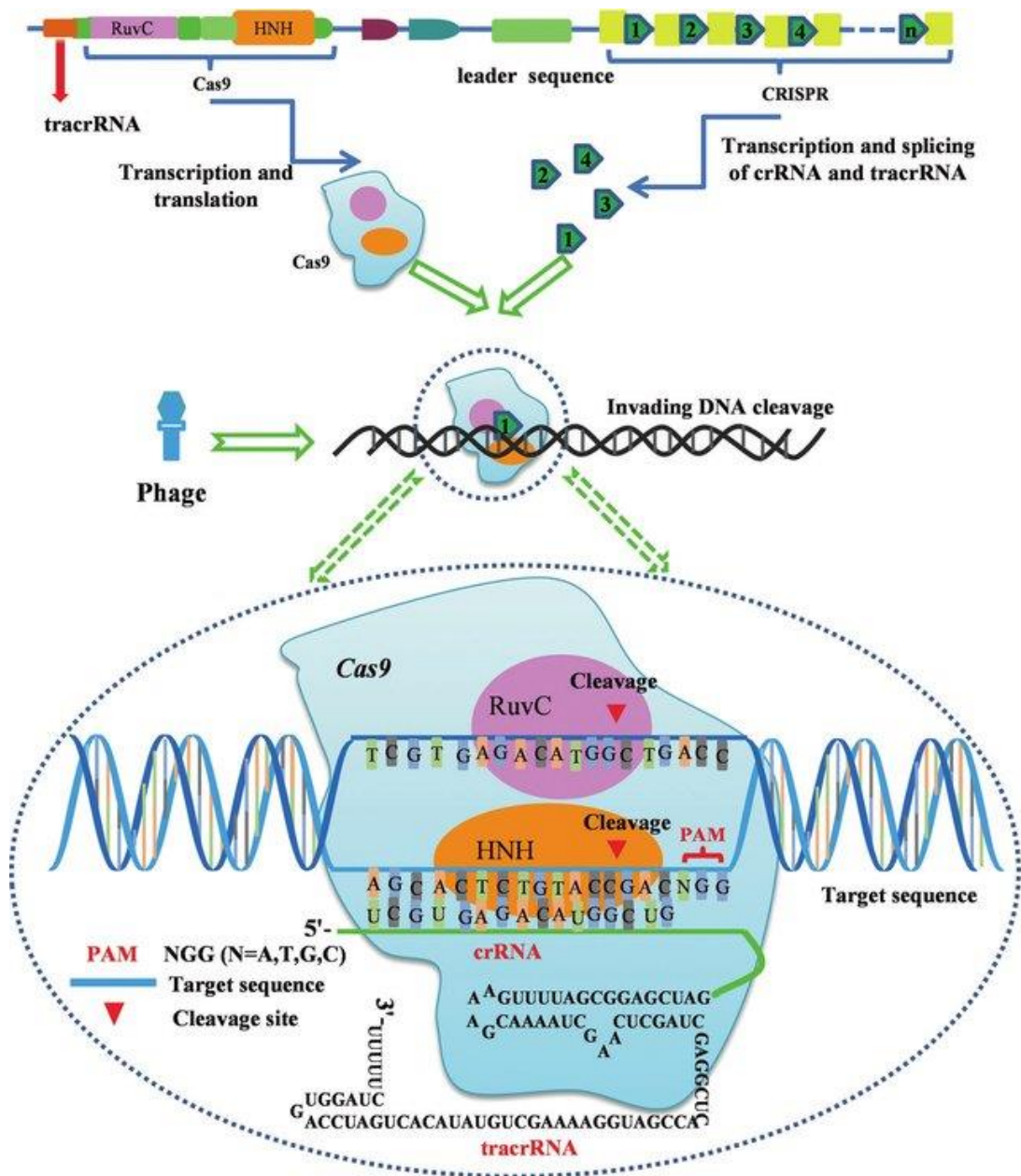


Figure 2.10: Molecular mechanism of CRISPR/Cas9-mediated DNA cleavage (Wang et al., 2018)

B. CRISPR/Cas 12

Cas12 also known as Cpf1 (CRISPR derived from *Prevotella* and *Francisella* 1). Unlike Cas9, Cas12 induced single strand break or stagger cut on target DNA sequence due to the presence of only 1 domain (known as RubC domain) unlike Cas9 that poses both RubC and HNH domains. In all the Cas effectors, Cas12 is the only Cas system that does not require

TracrRNA and recognize target sequence based on PAM sequence rich in Thymine (T) (Koonin et al., 2017).

C. CRISPR/Cas 13

Unlike CRISPR/Cas9, CRISPR/Cas 13 target and cleave RNA via 2 Higher Eukaryotic and Prokaryotic Nucleotide (HEPN) domains which exhibit RNase activity. Cas 13 enzyme form complex with mature CRISPR RNA and locate target RNA through PFS (protospacer flanking site) instead of PAM in Cas9 and cleave target RNA which activate the complex to cut other RNA as a result of a process known as “Collateral Activity”. Currently, there are so many applications of Cas13 for diagnostic and treatment of diseases based on RNA knockout and binding based on dCas13 which is the mutated version of Cas13 (i.e., inactive form) where the 2 HEPN domains are deactivated to bind to target without exerting collateral activity (Geraldi & Girl-Richman, 2018). The differences between all Cas effectors are presented in Table 2.5.

Table 2.5: Differences between Cas effectors

Differences	Cas 9	Cas 12	Cas 13
Domains	RubC and HNH	RubC	2 HEPN
Target	DSDNA	SSDNA	RNA
Organism derived from	Streptococcus pyogenes Streptococcus thermophilus Staphylococcus aureus	Prevotella sp. Francisella sp. Lachnospiraceae bacterium ND 2006 Acidaminococcus sp.	Prevotella sp. Leptotrichia wadei
Types of cut	Blunt	Staggard	-
TracrRNA	Present	Absent	Present
PAM sequence	NGG	T-rich	PFS

2.3.4 Application of CRISPR/Cas systems

Currently, there are so many applications of CRISPR/Cas systems undergoing clinical trials for editing diseases associated with human genes mutations such as treatment of sickle cell diseases and cancer therapies. Apart from gene editing using CRISPR/Cas systems to edit

human genes, scientist developed “base editing” which can edit DNA bases instead of genes. This approach has shown potentials for treatment of thousands of heredity diseases (Barrangou and Doudna, 2016)

2.3.5 CRISPR-based Biosensors

The field of biosensors is growing rapidly as a result of technological advancement in biology, electronics, computer science and nanotechnology. Biosensors are termed as analytical tools that integrate biological materials such as Nucleic Acid (NA), antibodies, enzymes, cell receptors etc integrated with transducers (such as electrochemical, magnetic, optical, piezoelectric, thermometric etc (Lazcka et al., 2007; Li et al., 2019).

Prior to integration of CRISPR/Cas systems in biosensors, scientist rely on benchwork diagnosis assays base on antibodies (recombinant polyclonal and monoclonal) and enzymes. These assays require trained technicians, long procedures, the use of reagents and pre-processing steps (such as extraction, amplification, centrifugation etc) (Nayak et al., 2009). other limitations associated with conventional biosensors include low sensitivity, accuracy and specificity (Stefano & Fernandez, 2017).

CRISPR-based biosensors have emerged as specific and sensitive molecular diagnostic tools for detection of pathogenic diseases, cancer mutations and genetic disorders. Unlike conventional biosensors, CRISPR-based biosensors are very specific due to hybridization with matching target sequence, rapid, sensitive, easy to use and cheap (Abudayyeh & Gootenberg, 2019) CRISPR-based biosensors are classified into 2 categories (binding and cleavage) according Cas effectors (Cas13, Cas12 and Cas9) (Li et al., 2019).

A) CRISPR/Cas9/dCas9-based Biosensors

Different CRISPR-based biosensors have been developed such the use of Cas9 and NA amplification based on Nucleic Acid Sequence Based Amplification (NASBA) or Nucleic Acid Sequence Based Amplification CRISPR Cleavage (NASBACC). These systems required amplification of target DNA which is cleaved by Cas9 resulting in signal generation (Li et al., 2019). The use of dCas9 along with single microring resonator has shown high sensitivity for detection of tick-borne disease which include Scrub typhus (ST) and Severe Fever Thrombocytopenia Syndrome (SFTS) (Koo et al., 2018).

Another CRISPR/Cas9 based biosensor was developed named CRISPR-Chip biosensor which utilize the binding activity of dCas9 and the high electron mobility of graphene-based effect transistor (gFET). However, the on-chip biosensor has shown high sensitivity for detection of genes associated with Duchene Muscular Dystrophy (DMD) (Hajian et al., 2019). In order to increase testing capacity of the pandemic caused by Zika virus, scientist developed paper-based optical CRISPR/Cas9 biosensor which can accurately discriminate between different viral strains at fM concentration (Pardee et al., 2016).

B) CRISPR/Cas12 based Biosensors

HOLMES (known as 1-Hour Low-cost Multipurpose Highly Efficient System) is one of the most popular Cas12a-based biosensors which detect target SSDNA quenched with fluorescent. The biosensor activity is based on the collateral cleavage activity of Cas12a isolated from *Lachnospiraceae* bacterium ND 2006 (LbCas12a) which form ternary complex with the target sequence resulting in cleavage of target sequence and trans-cleavage of non-target SSDNA as shown in Figure 2.11. HOLMES-based biosensor is versatile as it can function with or without amplification and can distinguish single base differences on target DNA sequence (Li et al., 2015). Cas12a-biosensor extracted from LbCas12a and *Acidaminococcus* sp. (AsCas12a) has shown pM sensitivity for detection of human papillomavirus (HPV-16) and Parvovirus (Batista & Pacheco, 2018).

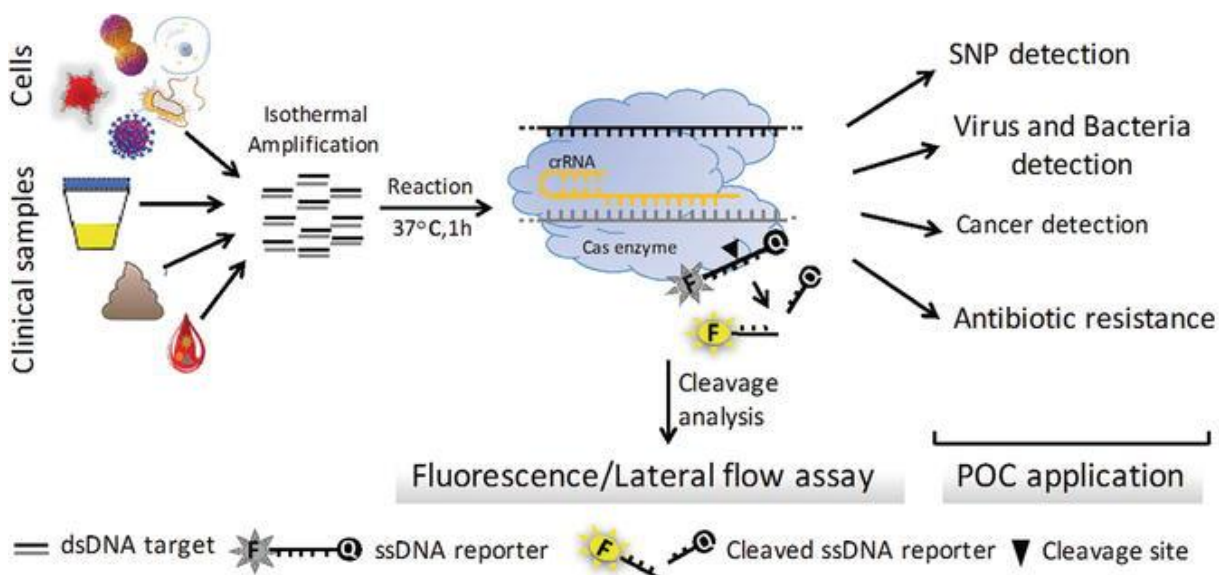


Figure 2.11: Cas12a/Cas13a-CRISPR-based Biosensor ((Geraldi & Girl-Richman, 2018)

C) CRISPR/Cas13 based Biosensors

CRISPR/Cas13-based biosensor is currently the most utilized CRISPR-based biosensor due to its collateral activity on reporter RNAs. An electrochemical CRISPR/Cas13-based biosensor has been developed to detect Micro RNAs (miR-lab) without amplification in blood serum of patients with brain cancer. The biosensor achieved detection limit in pM concentration (Bruch et al., 2019). SHERLOCK (Specific High-Sensitivity Enzymatic Reporter Unlocking) is another CRISPR/Cas13-based biosensor paired with Reverse Transcript Recombinase Polymerase Amplification (RT-RPA) or Recombinase polymerase Amplification (RPA) or and T7 transcription. The system utilizes Cas13a derived from LwCas13a for detecting both DNA and RNA target sequence at aM sensitivity for detection of Dengue virus and Zika virus, bacterial isolates, human DNA genotype, antibiotic resistance genes, cancer mutations etc. (Gootenberg et al., 2017) as shown in Figure 2.12.

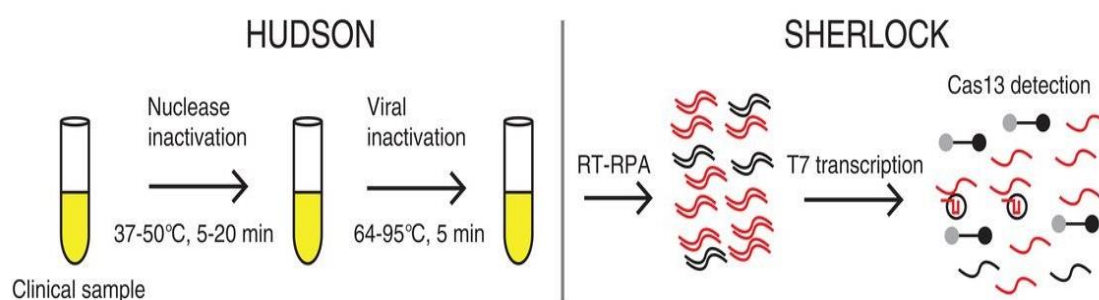


Figure 2.12: HUDSON and SHERLOCK

A second version of SHERLOCK V2 utilized Cas13b isolated from *Prevotella* sp. (PsmCas13b) instead of Cas13a in V1. The biosensor achieved higher sensitivity (2aM) than previous version for detection of gene mutations, Zika virus and Dengue virus in patient's liquid biopsy samples (Gootenberg et al., 2018). However, development of deployable biosensor has been a major challenge for scientist, recently, a group of scientists combine HUDSON (Heating Unextracted Diagnostic Sample to Obliterate Nuclease) with SHERLOCK for detection of viral DNA in bodily fluids (urine, serum and whole blood) within a sensitivity range of 20-90aM (Myhrvold et al., 2018).

2.3.6 CRISPR-based Biosensors for Detection of Tuberculosis

The use of CAS-EXPAR (CRISPR/Cas9 Exponential Amplification) which amplified target DNA for cleavage using Cas9 and the use of fluorescence monitoring has shown high

sensitivity for discriminating between single base mismatch at aM concentration (Haung et al., 2018). The use of 2 dCas9 enzymes in order to target amplified DNA sequence derived from *Mycobacterium tuberculosis* using PCR and detection method based on fluorescence tag has shown high detection sensitivity of equimolar sensitivity (Zhang et al., 2016).

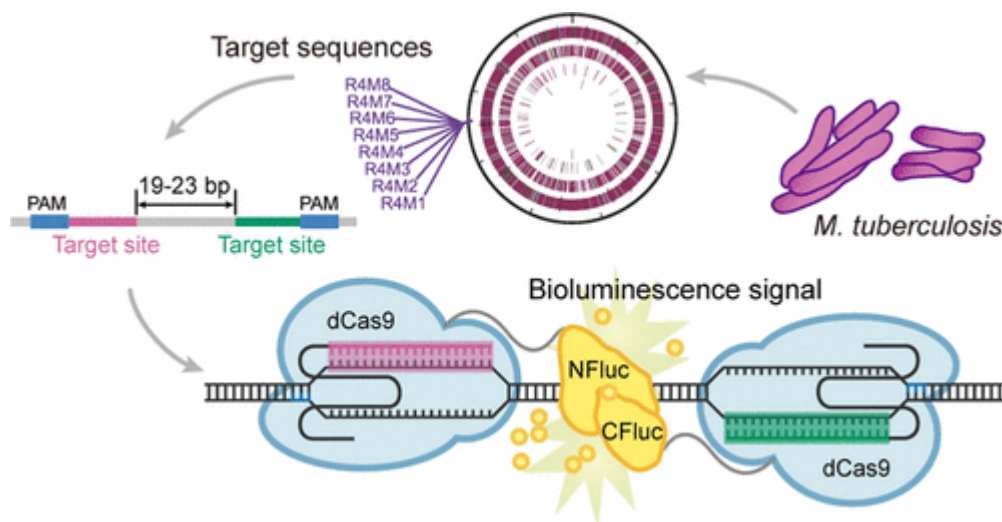


Figure 2.13: Detection of *Mycobacterium tuberculosis* using 2 dCas9

2.3.7 CRISPR-based Biosensors for Detection of COVID-19 Disease

The pandemic caused by SARS-CoV-2 has destabilized almost every sector leading to lockdowns, closure of borders, quarantine and massive screening. However, even with all these majors and restrictions, the number of infected patients continue to rise as well as death tools. In order to limit the spread of the virus through diagnosis, scientist rely on RT-PCR and qPCR method which is hindered by so many challenges. To addressed these challenges, researchers from Harvard and MIT developed a sensitive and specific paper based CRISPR/Cas13 biosensor for detection of SAR-CoV-2 RNA sequence. This technique is currently undergoing clinical trials and regulations (Patchsung et al., 2020; Zhang et al., 2020).

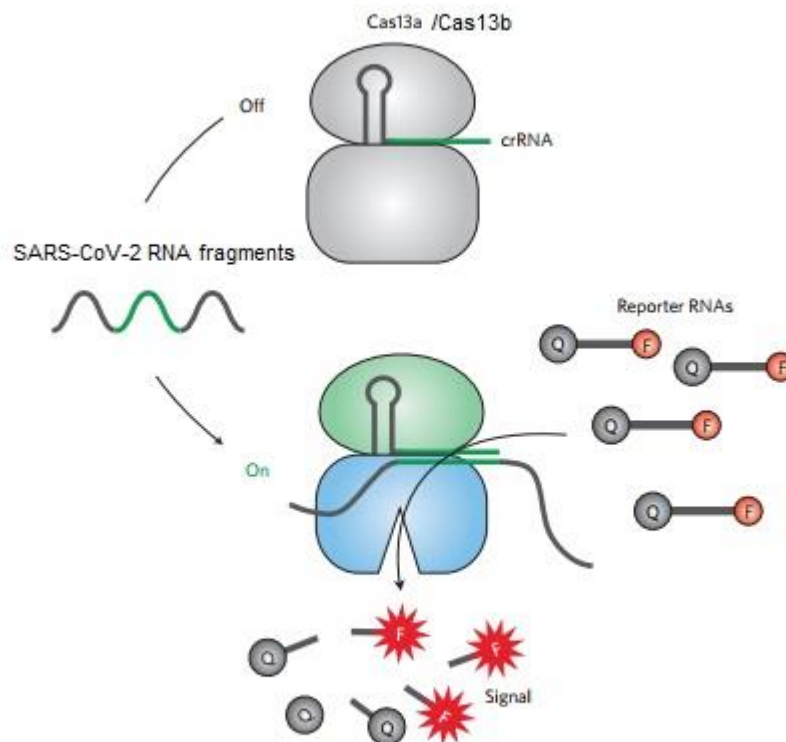


Figure 2.14: Detection of SARS-CoV-2 using Cas13a/Cas13b

These team of researchers employed SHERLOCK approach which has been successfully used for detection of Dengue virus, Zika virus and HPV (Myhrvold et al., 2018; Gootenberg et al., 2018). Unlike the current SHERLOCK technique, this new method revolves around the collateral cleavage activity of either Cas13a or Cas13b for detection of SARS-CoV-2 RNA fragments within 1 hour. The laboratory procedure for detection of COVID-19 rely on 3 steps (Zhang et al., 2020).

1. Sample collection and extraction: Nucleic acid are extracted from suspected patient's samples such as blood, nasal swab, urine etc.
2. Amplification: RPA technique is used for amplifying extracted viral RNA which last for 25 minutes.

3. Detection: Cas13a or Cas13b is employed which cleave target viral RNA fragments and the use of paper dipstick for visual read out which last for 32 minutes.

CHAPTER 3

LITERATURE REVIEW

Throughout last decade, scientists have been trying to integrate application of AI and its subfield which comprise of ML and DL in healthcare system. Researchers have utilized CNN to solve challenges in medicine such as detection using classification and segmentation approach for skin disease, brain cancer, breast cancer, diabetes (retinopathy) (Al-Turjman 2016; Serte & Demirel 2019). In the field of microbiology, Microbiologist, radiologist and computer scientist have been working side by side in order to detect microbial infections such as tuberculosis, malaria and pneumonia through the means of computer aided diagnosis (Kallianos et al 2019).

3.1 Tuberculosis

Automated detection of *Mycobacterium tuberculosis* has aided in accurate diagnosis of the disease. Smith et al., 2018 utilized a high-quality microscope designed to collect high resolution stain slide images. In order to make the bacilli visible, the researchers amplified the number of colonies and stained using dye to acquired 25000 images. Different types of image augmentation were carried out which are fed into a CNN model to discriminate various types of bacteria (rod, chain and round-shapes). The model achieved an overall accuracy of 95%. The classification of *Mycobacterium tuberculosis* and normal cases using ANN is provided by Khan et al., 2019. The study utilized over 12 thousand images which are partitioned into 70% for training of the model and 30% for validation. The image training was carried out using feedforward backpropagation model and the model achieved 94% testing accuracy.

To detect tuberculosis using deep learning approach such as image classification method, scientist utilized either microscopic slide image or chest X-ray radiographs. Lakhani & Sundaram (2017) employed CNN to discriminate between normal and pulmonary tuberculosis using radiographs. The study utilized 1007 posterior chest radiographs which were partitioned into training, validation and testing respectively and dataset was trained using

GoogleNet and AlexNet. The two models ensemble together to achieved AUC of 0.99 with a sensitivity of 97.3% and specificity of 94.7%.

A CNN Model built from scratch by Xiong et al., 2018 name TB-AI was used for detection of mycobacterium tuberculosis bacillus. The model was trained using 45 total samples with 15 as negative cases and 30 as positive cases which are tissue samples that were treated using acid-fast stain. The result has shown TB-AI achieved 83.65% specificity and 97.94% sensitivity. A study based on model designed from scratch is proposed by Aguiar et al., 2016. The study was carried out to classify pulmonary TB using ANN designed using multi-perceptron. The model was trained using different types of datasets such as radiological images, pulmonary TB cases acquired from bronchoalveolar lavage and sputum samples (with 315 total cases), respiratory symptoms and patient's demographics. The dataset was partitioned into 80% and 20% for training and validation respectively and the model achieved 89% specificity and 96% sensitivity.

Panicker et al., 2018 utilized CNN approach to detect Mycobacterium tuberculosis bacillus from microscopic sputum smear images. The dataset was obtained from a public domain with 120 images which were cropped to 900 patches for both positive and negative samples for segmentation method. The model accomplished sensitivity of 97.13% and specificity of 78.4%. Quinn et al., 2016 applied deep learning approach to diagnose tuberculosis from sputum smear sample stained with Ziehl Nelson stain. The authors used microscope to obtained 315,142 test patches. The results have shown a higher AUC value of 0.99 and the potential of the CNN model in classifying test samples.

A study based on the use of SVM for detection of Mycobacterium tuberculosis is provided by Costa et al., 2015. The study employed 120 smear microscopic slide image from 12 cases. Prior to training the images undergo conventional smear microscopy and segmentation. The study reported an error rate of 3.38% and sensitivity of 96.8%. The same approach was adopted by Yahiaoui et al., 2017 to classify tuberculosis from healthy samples using SVM. The model was trained using 15 total CXR images (which are acquired from 50 patients suffering from TB and 100 healthy samples). However, the model achieved 96.7% accuracy. A deeper model was employed by El-melegy et al., 2019 for detection of Ziehl Nelson stained sputum smear of tuberculosis and healthy images. The research utilized 500 images which are divided into 80% and 20% for training and validation respectively and train using

Faster Region-based convolutional neural network plus CNN (F-R-CNN+CNN) and Region-based convolutional neural network F-R-CNN. F-R-CNN+CNN achieved 85.1% sensitivity and 98.4% accuracy while F-R-CNN achieved 82.6% sensitivity and 98.3% accuracy.

The use of TL based on VGGNet and SVM as the model classifier is reported by Ahsan et al., 2016. The study utilized dataset obtained from Shenzhen hospital, China and from Montgomery County Tuberculosis Control Program (MCTCP) in the form of CXR images. The dataset was trained based on (I) with augmentation and (II) without augmentation. The model achieved 81.25% validation accuracy with augmentation and 80% validation accuracy without augmentation. Chang et al., 2020 proposed a 2-stage classification of tuberculosis based on TL on CNN. The study made use of 1727 cases of tuberculosis culture acquired from Tao-Yuan general hospital, Taiwan. The dataset was trained using VGGNet, YOLO and CNN designed from scratch. By targeting the result of the model on non-negative class, the proposed system achieved 98% recall and 99% precision.

Due to the prevalence of tuberculosis in Uganda, Muyama et al., 2020 utilized 3 TL models based on ResNet (inception V3), GoogleNet and VGGNet for computer assisted-detection of tuberculosis from Ziehl-Nelson sputum smear slide images. The study made use of dataset obtained from an online database and the ones captured using cell phone's camera in the university microbiology laboratory. The 2 datasets are combined together and partitioned into 80% for training and 30% for validation. The models were trained according to (I) find-tuning (II) with augmentation and (III) without augmentation. However, among all the pre-trained models, ResNet achieved the highest accuracy score of 86.7%.

To classify normal and abnormal X-ray images of individuals suffering from tuberculosis, Abbas and Abdelsamea, 2018 utilized a pretrained AlexNet model. The model was trained based on 138 total CXR images (58 normal and 80 abnormal X-ray images). To increase number of training set, the X-ray images undergo data augmentation. The model hyperparameters were turned according to deep-tuning, shallow-tuning and fine-tuning techniques. The study revealed that hyperparametric fine-tuning of pretrained AlexNet outperform other tuning techniques with 0.998 AUC score, 99.7% sensitivity and 99.9% specificity.

The application of hybrid model for automated detection of tuberculosis is proposed by Sahol et al., 2020. The hybrid model comprises of MobileNet with 88 layers and feature selector in the form of Artificial Ecosystem-based optimization (AEO) algorithm. The model was trained on X-ray dataset acquired from Shenzhen hospital, China with 662 totals frontal CXR images (of which 336 are positive and 336 negative). The model achieved 90.2% best classification accuracy, 93.85% sensitivity and 86.76% specificity.

Table 3.1: Detection of tuberculosis using AI-driven tools

Reference	Neural Network	Sample type	Data set/Training	Results
Khan et al., 2019	CNN	Microscopic Stained image	12,636	94% accuracy
Costa et al., 2015	SVM & CNN	Microscopic stained images	120	96.80% accuracy
Smith et al., 2018	CNN	Microscopic stained image	25,000 images	95% accuracy
El-Melegy et al., 2019	CNN	Microscopic stained image	500 images	F-R-CNN achieved 98.3% accuracy and 82.6% sensitivity F-R-CNN+CNN achieved 98.4% accuracy and 85.1% sensitivity
Xiong et al., 2018	CNN	Microscopic stained image	Cases: 45 30 positives 15 negatives	97% sensitive and 83.65% specific

Klassen et al., 2018	Automated recognition and pattern analysis	Flouographic chest Images	Negative = 238 Positive = 70	sensitivity 75.0- 87.2%, specificity 53.5- 60.0%,
Aguiar et al., 2016	ANN (multi- perceptron)	radiological data and pulmonary tuberculosis sputum samples	315 cases, 80% training and 20% testing	96% sensitivity and 89% specificity
Yahiaoui, & Yumusak, 2017	SVM	chest x-ray images	150 cases 50 positive and 100 negatives	96.68% accuracy
Muyama et al., 2020	ResNet GoogleNet VGGNet	Ziehl-Nelson- stained smear sputum slides	Not specified	86.7% accuracy
Abbas & Abdelsamea, 2018	Pretrained AlexNet model	Chest Xray	138 (58 negative and 80 positive cases)	0.998 AUC score, 99.7% sensitivity and 99.9% specificity
Sahol et al., 2020	Hybrid model (MobileNet and Artificial Ecosystem- based optimization algorithm	Frontal Chest Xray	662 (336 positive and 336 negative)	90.2% best classification accuracy, 93.85% sensitivity and 86.76% specificity.

2.2 Pneumonia

Chest radiographs based on Computed Tomography (CT) scan or Chest X-ray (CXR) is a technique employed by radiologist to discriminate between individual affected with pneumonia and healthy (i.e., normal) ones. The major difference between Xray images of infected person and healthy ones due to the appearance of white hazy spots which also described as “Ground-glass opacity” in individual positive of pneumonia which is absent in healthy individuals.

Xray images are the basic components used for detection of pneumonia using Machine learning approach. This idea is adopted by Stephen et al., 2019. The authors employed DL method to classify samples of CXR images. The research employed a CNN that is built from scratch using Keras open source with TensorFlow to extract distinctive features from positive and negative images. The dataset contain 5856 X-ray images of normal and pneumonia images collected from pediatric patients between 1-5 years old. The dataset was further augmented to yield a greater number of training dataset. The model was tested on different data size (100-300) and the model achieved average accuracy of 94.81% 93.01% training and validation respectively.

ChestX-Ray8 is a new dataset from Chest X-Ray Database and Benchmarks which is utilized by Wang et al., 2019. The datasets contain X-ray images with total number of 108,948 from 32,717 patients for detection of thoracic diseases. The authors utilized the dataset and trained using CNN networks such as AlexNet, VGGNet-16, GoogleNet and ResNet-50. The research achieved AUC value of 0.6333 for “Pneumonia. A similar study adopted by Rajpurkar et al., 2017 who developed a 121 CNN called CHeXNet. The research utilized more than 100 thousand frontal view CXR images with more than 10 diseases. For detection of Pneumonia, the model achieved AUC of 0.8887 with the model outperforming Radiologist.

Saraiva et al., 2019 Classified x-ray Images of Children positive of Pneumonia using CNN models. The research dataset was obtained online from research of Kermany et al., 2018 characterized as Optical Coherence Tomography (OCT) and CXR Images with total number of 5863 images. The model was train base of cross validation ($k = 5$) and the model achieved 95.30% average accuracy. Recently, Chouhan et al., 2020 adopted TL approach to classify X-ray images into positive and negative pneumonia samples. The research employed transfer

learning models of Resnet (Inception V3), GoogleNet, DenseNet121 and AlexNet. A total of 5856 normal and pneumonia (bacteria and virus) were used. The model achieved respective training (at different epochs for the models) and testing accuracies with AlexNet (98.97 and 92.86%), DenseNet121 (99.23 and 92.62%), GoogleNet (99.48 and 93.12%) and ResNet (99.48 and 94.23%).

To discriminate between viral and bacterial pneumonia, Rajaraman et al., 2018 employed CNN (VGG-16, Residual and Inception CNN) for detection of pneumonia in Pediatric Chest Radiographs by localizing the region of Interest (ROI). The dataset contains total number of 5856 including normal CXR images, bacterial pneumonia and viral pneumonia CXR images. The models achieved 96.2% for bacterial pneumonia and 93.6% for viral pneumonia. A more sophisticated study is carried out by Zech et al., 2018 who utilized Deep NN and split validation approach to screen for pneumonia in CXR images. The researched employed a total number of 158,323 CXR images collected from 3 different platforms. The results have shown higher accuracy and AUC values with distinction of the 3 institutions where the dataset was acquired.

As the need for accurate detection of pneumonia in underdeveloped countries continue to rise due to unavailability of sensitive testing kit, Rahman et al., 2020 proposed the use of 4 different pretrained CNN models (SqueezeNet, AlexNet, DenseNet and ResNet18). The models were trained using 5247 total CXR images which contain 2561 bacterial pneumonia, 1345 viral pneumonia and 1341 healthy CXR images which are made available online (Kaggle website). Dataset was split into training and validation and the training dataset undergoes augmentation to increase the number of images which are fed into the models. The outperformance on test set has shown that the models achieved 98% accuracy for classification of healthy cases vs pneumonia, 95% for classification of bacterial vs viral pneumonia and 93.3% for classification of healthy CXR images, viral pneumonia and bacterial pneumonia.

The application of ensemble CNN has shown great promise for classification of images. A study conducted by Liz et al., 2020 ensembled CNN and design of Explainable Artificial Intelligence (EAI) (a novel clinical support system) for the classification of pneumonia into consolidation and non-consolidation base on heatmaps. The study made use of 950 X-ray

images for training and validation which result in 0.92 AUC score and 0.73 True Positive Rate (TPR) score.

Table 3.2: Table 2: Detection of different types of pneumonia using AI-driven tools.

Reference	Type of pneumonia	Dataset	Result
Stephen et al. 2019	Viral pneumonia (Influenza virus)	5856 CXR images	94.81% average training accuracy and 93.01% validation accuracy
Rajpurkar et al. 2017	Not specified	108,948 X-ray images	0.6333 AUC score
Wang et al. 2017	Not specified	100, 000 X-ray images	0.8887 AUC score
Wang et al. 2020	Viral pneumonia (COVID-19)	453 CT scan images	82.9% validation accuracy, 84% sensitivity and 80.5% specificity. 73.1% testing accuracy, 74% sensitivity and 67% specificity
Saravia et al. 2019	Viral pneumonia (strain not specified)	5863 CXR Images	Accuracy of 95.30%
Chouhan et al. 2020	Viral and Bacterial pneumonia (strains not specified)	5863 Chest X-Ray Images	Different models were used
Xu et al. 2020	viral pneumonia (COVID-19, Influenza-A)	618 CT scan Images	Accuracy of 86.7%.
Rajaraman et al. 2018	Viral and Bacterial pneumonia (strains not specified)	5856 CXR images	96.2% accuracy for bacterial pneumonia and 93.6% for viral pneumonia
Zech et al. 2018	Viral and Bacterial pneumonia (strains not specified)	158,323 chest radiographs	Different models were used
Liz et al., 2020	Viral Pneumonia	950 X-ray	0.92 AUC score and 0.73 True Positive Rate (TPR) score

3.3 COVID 19

3.3.2 COVID 19 vs Normal CXR images

The use of 3 CNN models which consist of models built from scratch and TL models (Inception-ResNetV2, Inception V3 and pretrained ResNet-50) for screening of COVID-19 is proposed by Narin et al., 2020. The research utilized 100 total CXR images (50 COVID-19 cases and 50 normal images). The dataset was partitioned into 80% and 20% for training and validation respectively. The models were training based on the data split and 5-fold cross validation. The models achieved 98% accuracy for pretrained ResNet50, 87% accuracy for Inception-ResNet V2 and 97% accuracy for Inception V3.

Mei et al., 2020 proposed the use of AI-driven networks to classify COVID-19 based on CT scan images of patients diagnosed using laboratory testing kit (RT-PCR and next-generation sequencing RT-PCR) along with exposure history and clinical symptoms. The study utilized 905 total number of patients which are partitioned into 626 for training and 279 for testing. The dataset was trained using Deep CNN which learn image characteristics and the use of 3 classifiers (multilayer perceptron, SVM and random forest) for classification. The model performance was compared with senior thoracic radiologist and the model achieved AUC value of 0.92 outperforming the radiologist in terms of total positive rate.

The use of lightweight Xray images obtained from mobile phone for screening of COVID-19 using CNN is proposed by Zulkifley et al., 2020. The model named LightCovidNet was trained on different Xray images achieving 92.37% accuracy. Turkoglu (2020) proposed the use of DL model known as COVIDetectionNet model, a pretrained AlexNet model integrated with SVM as a classifier, the model was trained using 6092 total CXR images which contain both COVID-19 pneumonia and normal images and it attained an AUC value of 99.76 and accuracy of 98.43% on testing dataset.

Another study which makes use of TL is proposed by Mohammedi et al., 2020. Different pretrained models (MobileNet, VGG-19, VGG-16 and ResNet/Inception V2) are used for binary classification of COVID-19 from healthy CXR images. The study employed 348 total CXR images (112 positive and 236 negative). To increase the number of datasets, data augmentation techniques were carried out by zooming, rotation, flipping (horizontal and vertical), height shift, width shift and filling (nearest mode). After training the models were validated on unseen dataset with MobileNet achieving highest classification performance resulting in 99% F1 score, 99.1% accuracy, 100% precision and 98% recall.

Table 3.3: Binary classification of CXR images using AI-Driven Models

Reference	Model type	Dataset	Result
Narin et al., 2020	COVID-19	100 total images (50% COVID-19 and 50% healthy CXR images)	87% accuracy for Inception-ResNetV2 and 97% accuracy for InceptionV3
Abbas et al., 2020	COVID-19	185 normal CXR images and 11 COVID-19	The model achieved 95.12% accuracy, 97.91% sensitivity of 97.91% and 91.87% specificity of 91.87%.
Wang S et al., 2020	COVID-19	453 COVID-19 CXR images	The model achieved external 73.1% testing accuracy, 74% sensitivity and 67% specificity
Mohammedi et al., 2020	MobileNet, VGG-19, VGG-16 and Inception V2	348 total CXR images (112 positive and 236 negative).	MobileNet achieved 99% F1 score, 99.1% accuracy, 100% precision and 98% recall
Zulkifley et al., 2020	CNN	-	92.37% accuracy
Turkoglu (2020)	COVIDetectionNet plus SVM	6092 total CXR images	The model achieved an AUC value of 99.76 and 98.43% accuracy on testing dataset.
Mei et al., 2020	Deep CNN plus classifiers (multilayer perceptron, SVM and random forest)	626 CT scans	AUC value of 0.92
Ozturk et al., 2020	DarkNet model	-	The model achieved 98.08%

3.3.2 Multiclass (Covid 19, Non-Viral COVID-19)

The use of DL system to screen and classify the novel 2019 coronavirus and influenza-A viral pneumonia using CT scans was proposed by Butt et al., 2020, the study acquired 618 total transverse section CT scans with 219 images from 110 patients suffering from COVID-

19 disease (confirmed by RT-PCR testing method), 224 CT scans images of patients suffering from Influenza-A viral pneumonia and 175 CT scans images of control (normal cases). The model performance resulted in 0.996 AUC score for discrimination of COVID-19 vs non-COVID-19 viral pneumonia with 98.2% sensitivity and 92.2% specificity.

In order to classify non-COVID-19 viral pneumonia, COVID-19 pneumonia and bacterial pneumonia using CXR images. Hall et al., 2020 utilized pretrained DL neural network to trained the model using over than 4 thousand bacterial pneumonia and non-COVID-19 viral pneumonia and 124 COVID-19 CXR images. The model performance resulted in an AUC value of 0.997 and 95% accuracy for discrimination of other viral pneumonia (bacterial and non-COVID-19) and 100% accuracy for COVID-19 on unseen (testing dataset).

Truncated Inception ResNet based on TL is utilized by Das et al., 2020 to scan different classes of pneumonia (Non-COVID-19 and COVID-19 cases), tuberculosis and normal CXR cases. The study utilized 162 COVID-19 CXR images granted by Cohen et al., 2020, 4280 viral pneumonia and 1583 Normal/healthy CXR images from Kaggle and 682 CXR cases from Montgomery county and Shenzhen China. The model achieved AUC value of 1.0 and 99.96% accuracy for classification of COVID-19 from both healthy CXR images and viral pneumonia. For multiclass (COVID-19, tuberculosis, viral pneumonia and normal CXR images), the performance of the model resulted in an AUC value of 0.99 and 99.92% accuracy.

Ozturk et al., 2020 proposed the use of automated approach for the screening of COVID-19 from normal CXR images. The dataset used for the study was made available by Cohen et al., 2020. The images were trained using DarkNet model designed using 17 convolutional layers to classify binary dataset (COVID-19 cases and No findings) and multiclass (COVID-19 cases, other viral pneumonia and No findings). The model performance resulted in 98.08% for binary classification and 87.02% for multiclass classification.

The application of TL (pretrained models) on DL for classification of multi-types of pneumonia is proposed by Chowdhury et al., 2020. The study utilized dataset made available online to create a single public database that combined all the dataset. The model was trained using 423 CXR images COVID-19 pneumonia, 1579 healthy CXR images, 1485 non-COVID-19 viral pneumonia. The performance evaluation has demonstrated that models achieved 99.7% sensitivity and 99.7% specificity for discrimination of COVID-19 and

normal CXR images with augmentation and 99.7% sensitivity and 99.55% specificity for discrimination of COVID-19 and normal CXR images without augmentation. In terms of multiclass, the models achieved 97.9% sensitivity and 97.95% specificity for classification of COVID-19, non-COVID-19 viral pneumonia and healthy CXR images with augmentation and 97.9% sensitivity and 98.885% specificity for classification of normal CXR images, Non-COVID-19 viral pneumonia and COVID-19 pneumonia without augmentation.

A combination of 2 models (i.e., hybrid networks) for 2-way screening of positive COVID-19 and negative cases based on CXR images and 3-way screening of viral pneumonia, bacterial pneumonia and COVID-19 pneumonia was stated by Quan et al., 2020. Detection of COVID-19 was made possible by fusion of CNN models in the form of DensNet and capsule network known as CapsNet on dataset containing 1472 CXR images for binary discrimination of COVID-19 and normal CXR images. The multiclass classification utilized dataset made available by Linda Wang et al., 2020 which contain closed to 14 thousand X-ray images. The model performance resulted in 99.32% validation accuracy, 100% specificity and 98.85% sensitivity.

The use of CNN for classification of normal CXR images, COVID-19 pneumonia, non-COVID-19 viral pneumonia was reported by Echtioui et al., 2020. The study utilized 500 COVID-19 CXR images acquired from GitHub repository (an open-source dataset) and 1000 (500 non-COVID-19 viral pneumonia and 500 normal CXR images acquired from ChestXray8 from the same repository. The model was trained based on CNN model with 13 total layers (10 convolutional layers and 3 fully connected layers. The validation of the model on unseen dataset resulted in 91.34% accuracy for 3-way discrimination Normal CXR images, COVID-19, non-COVID-19 viral pneumonia.

Table 3.4: Multiclass detection of Pneumonia using AI-driven tools.

Reference	Type of pneumonia	Dataset	Result
Li et 2020 al.,	COVID-19 and Community Acquired Pneumonia (CAP)	4352 CT scans (1292 of COVID- 19, 1735 of CAP and 1325 normal CT scans)	The models achieved 90% *SV and 96% *SF for detection of COVID-19 and 87% *SV and 92% *SF for detection of CAP

Chowdhury et al., 2020	COVID-19 and non-COVID-19 VP	423 COVID-19, 1458 viral pneumonia and 1579 normal Chest X-ray images	The models performance resulted in higher accuracies, sensitivities and specificities
Mahmud et al., 2020	COVID-19, non-COVID-19 VP, BP	1493 non-COVID-19 viral pneumonia, 305 COVID-19 pneumonia, 2780 bacterial pneumonia	The model performance resulted in 97.4% *AC for COVID-19 vs normal, 96.9% for COVID-19 Vs non-COVID-19 VP, 94.7% for COVID-19 vs BP and 90% for multi-class
Rajaraman et al., 2018	Non-COVID-19 VP and BP (strains not specified)	5856 chest X-ray	The model achieved *Ac of 96.2% accuracy for BP and 93.6% for non-COVID-19 VP
Quan et al., 2020	COVID-19, non-COVID-19 viral pneumonia and normal X-ray images	14 thousand CXR images.	The model performance resulted in 99.32% validation accuracy, 100% specificity and 98.85% sensitivity.
Echtioui et al., 2020	COVID-19, non-COVID-19 viral pneumonia and normal X-ray images	500 COVID-19, 500 non-COVID-19 viral pneumonia and 500 normal X-ray images	The model achieved 91.34% accuracy
Ozturk et al., 2020	COVID-19 cases, other viral pneumonia and No findings		The model achieved 87.02% accuracy

CHAPTER 4

EXPERIMENTAL SET UP

The overall method behind this research is shown in flow chart in figure 5.1. The method can be summarized based on data collection from different sources such a Near East Hospital and dataset made available by online website such as GitHub and Kaggle, image processing is conducted for the purpose of reducing the size of the images to fit into AlexNet model. Other approaches employed include Data augmentation which is carried out on training dataset, model training using Matlab and performance evaluation based on accuracy, sensitivity and specificity.

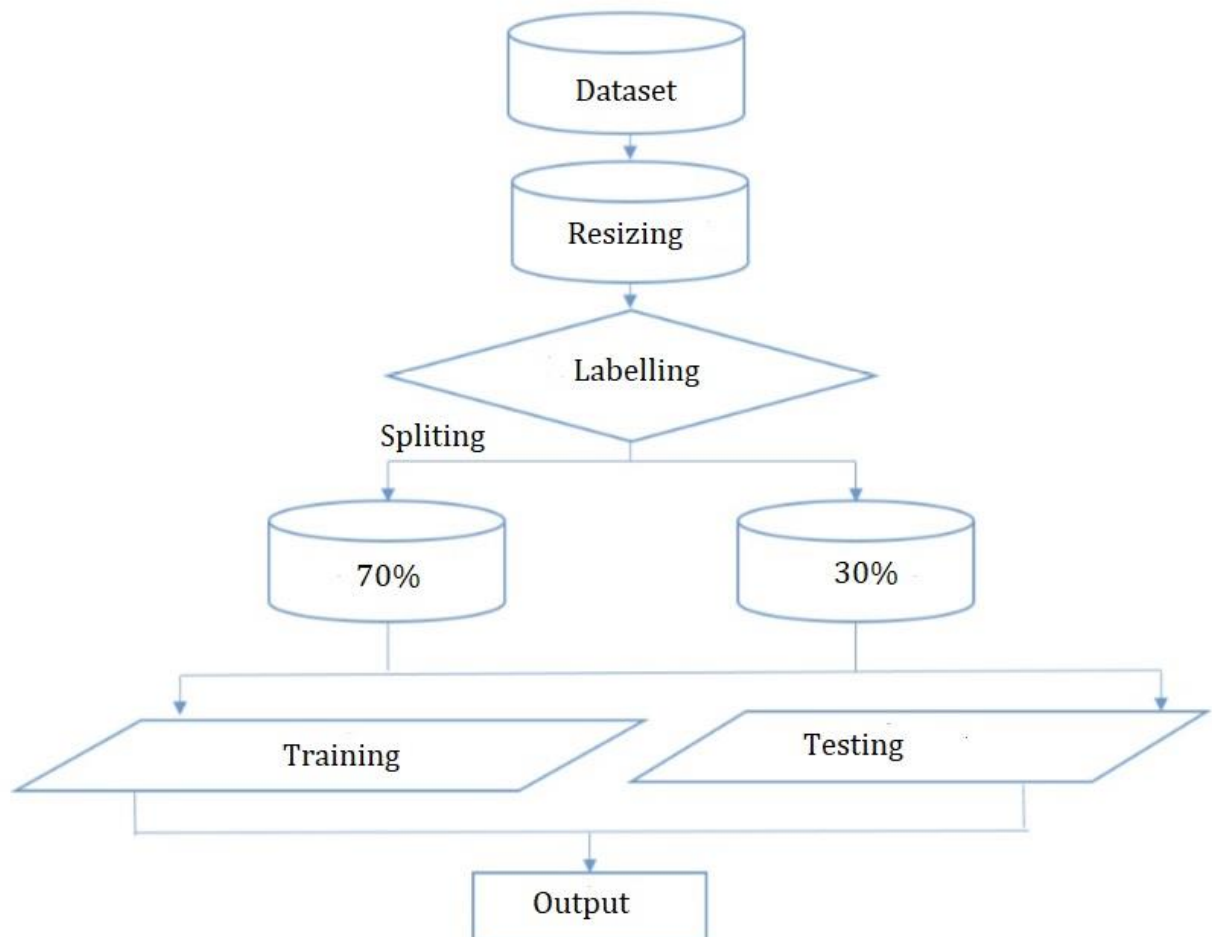


Figure 4.1: Work flow chart

In this study, 2 different types of dataset are used; Microscopic slide images for classification of tuberculosis and chest X-ray images for classification of pneumonia and its sub-class

which include COVID-19 pneumonia, Non-COVID-19 viral pneumonia (e.g., Influenza virus) and bacterial pneumonia.

4.1. Tuberculosis

4.1.1 Data collection

The sputum smear slides utilized in this experiment were prepared by smearing the sputum specimen on a clean slide. A low flame was used to fix the air-dried smear for 2-3 times. Auramine O stain is then utilized to fill the slides in order to bind on to the cell wall of the acid-fast bacteria which last for 10 minutes. subsequently, decolorization technique was carried on the slides using alcohol and later washed using running water. Potassium permanganate was employed to counter stained the slides in order to obtain a clear contrast background prior to rinsing with clean water and dried using air. (Priya & Srinivasan 2016). The images were acquired using a camera in monochrome binning mode attached to a 20× objective fluorescence microscope of 0.5 numerical aperture. The camera (AxioCam HR) has a resolution of 4164×3120 with a pixel size $6.45 \mu\text{m (h)} \times 6.45 \mu\text{m (v)}$.

Positive cases (Ziehl Nelson acid fast stained) were obtained from Near East University Hospital plus 100 from Istanbul Tuberculosis Control Association (ITCA). Negative samples were prepared by pathologist using the same staining approach on samples obtained from individuals without suspicion of tuberculosis leading to 530 total images as shown in Figure 4.2

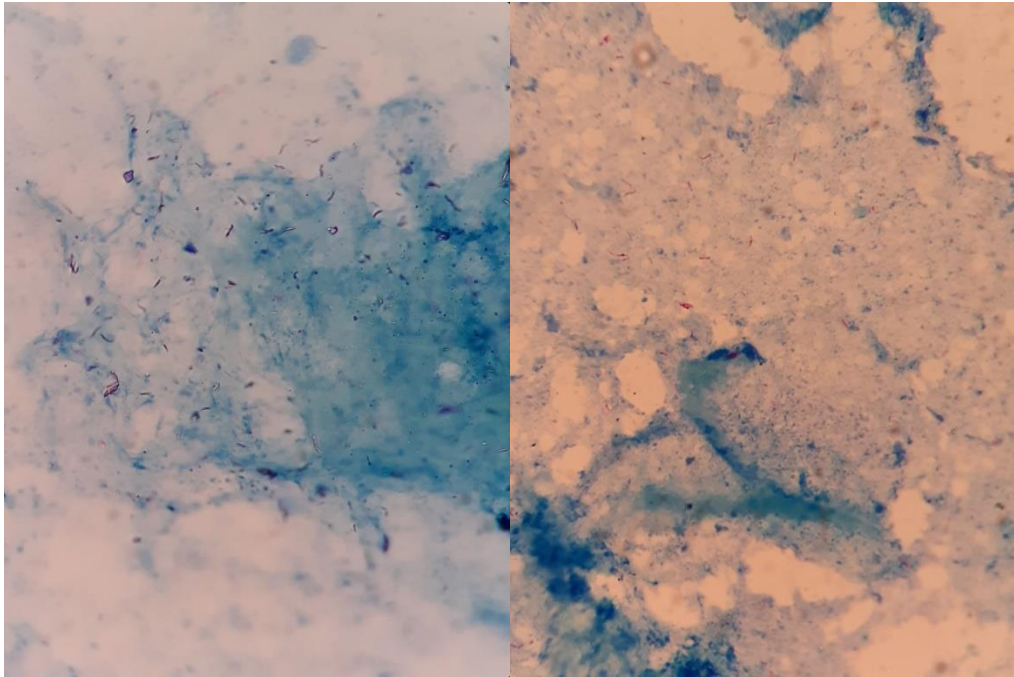


Figure 4.2: Positive (tuberculosis) microscopic slide image. The purple and red thick bacilli depict mycobacterium tuberculosis

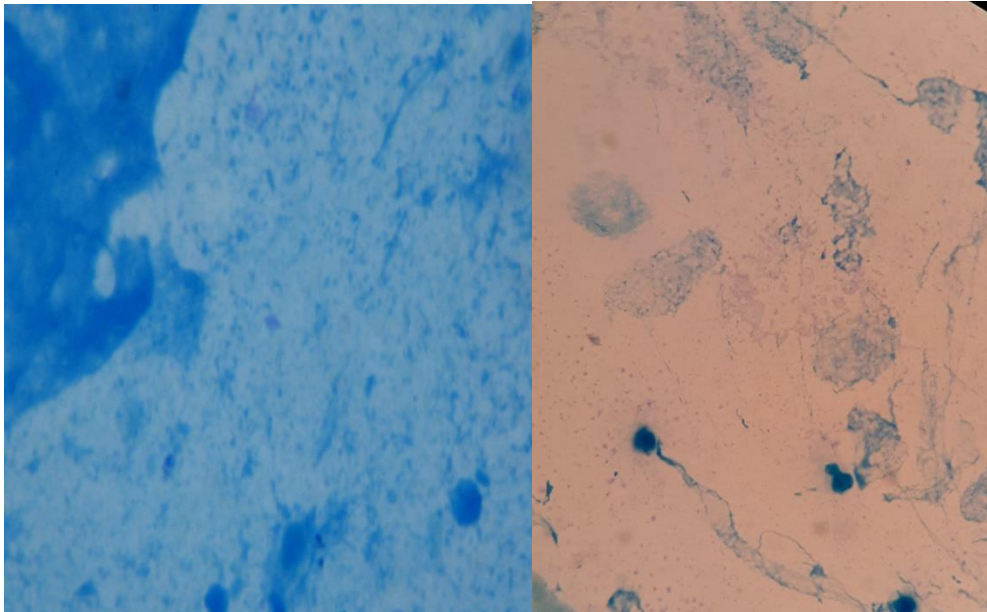


Figure 4.3: Negative (tuberculosis) microscopic slide image

4.1.2 Data Processing

All the images acquired are evaluated by specialist by going through each image and labelling them accordingly. For Tuberculosis, the images are labelled into 2 groups, positive

(due to the presence of red bacilli after staining) and negative (due to the absence of red bacilli after staining) as shown in figure 4.2 and 4.3. For Chest X-ray images of pneumonia, images are labelled into 4 groups, COVID-19 pneumonia, non-COVID-19 viral pneumonia, bacterial pneumonia (due to the presence of multilobar ground glass opacities with posterior or peripheral distributions) and normal (i.e., healthy, negative due to the absence of multilobar ground glass opacities) as shown in figure 4.5 and 4.7.

The images acquired ranges from 2 mega bite (MB) to 4MB, with pixels sizes (based on horizontal and vertical) ranging from 2000X3500 to 3000X5000. These images are too big to fit into AlexNet Models. However, in order to reduce them to 227X227X3 pixels size, an online system was employed known as BIRME available at this website: https://www.birme.net/?target_width=227&target_height=227

The interfaced of the settings is shown in Figure 4.4.

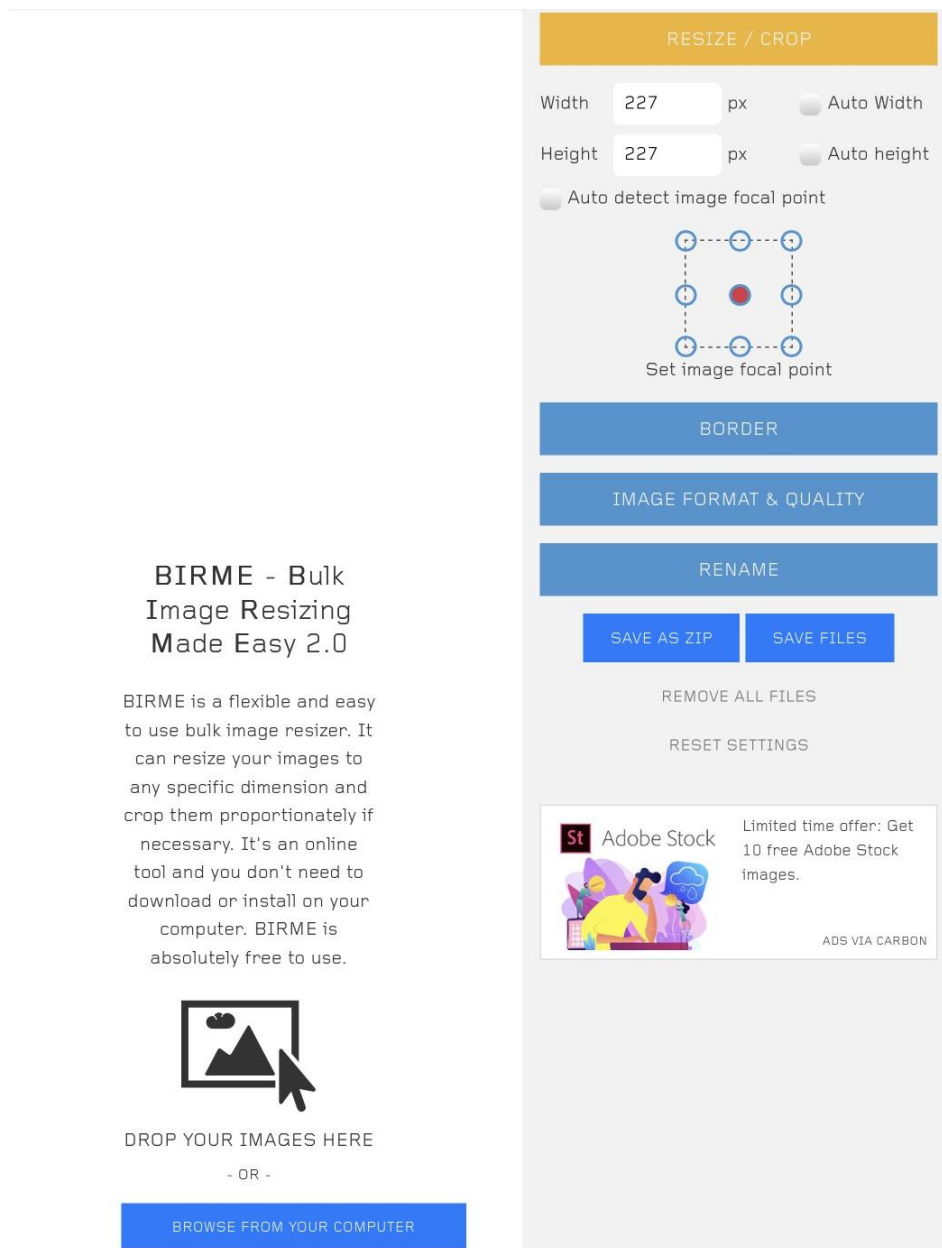


Figure 4.4: BIRME Interfaced

After resizing the images to the required pixels size and labelling each type according to respective classes, the images are fed into the Matlab using the codes presented in Appendix A.

4.1.2 Data split

Splitting dataset is integral for training and validating performance of AI-driven models. Scientist recommends the use of different data split such as 70:30, 75:25 and 80:20.

However, we employed 70:30 split in order to have enough unseen dataset and for generalization. Based on the 530 total microscopic slide images validated by pathologist, 367 are used as training (70%) and 159 images as testing (30%) dataset.

4.1.3 Data augmentation

Data augmentation is only carried out using training dataset mainly to increase the number of images for the model to learn efficiently. The augmentation approaches employed include cropping and rotation (90^0 , 180^0 and 270^0) on 371 images resulting in 2444 total training dataset in which 1320 are positive (tuberculosis) and 1144 are negative (healthy). Prior to data augmentation, the slide images are more than 3456x4608x3 pixels size. Size reduction is carried out to reduce the images to 227x227x3 to fit into pretrained AlexNet model.

Table 4.1: Data augmentation of Training Dataset

Label	Training
Positive	1320
Negative	1144
Total	2464

4.1.4 Machine Vs Human

In order to compare the learning competence of the model on unseen dataset with ground truth (which are real samples validated by certified pathologist), we tested the model using 30 unseen images. The same number of images are also used to compare the model with both beginners and certified pathologist.

4.2. Pneumonia

This study employed dataset (based on X-ray images) which are uploaded by Kermany et al., 2018. The images are categorized in to 3 main folders namely, training, testing and validation. The sum of the folders resulted in 5856 positive and negative cases as shown in Table 4.2. The portrait of the dataset is according to radiographic images acquired from patients older than 1 year and below 5 years as shown in Figure 4.3.

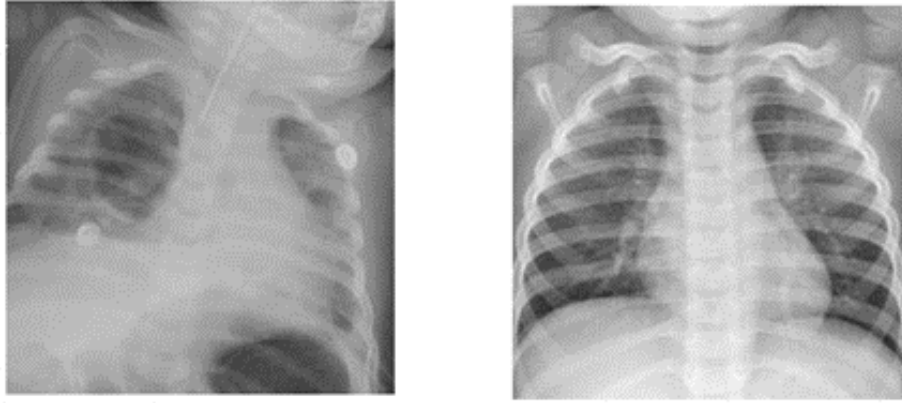


Figure 4.3: Pediatric CXR scans. Left: Pneumonia. Right: Normal CXR scan.

Table 4.2: Datasets Description

Label	Number
Positive	4273
Negative	1583
Total	5856

4.2.2 Data split

The datasets are partitioned into 50:50, 60:40, 70:30, 80:20 and 90:10 for both model training and model validation respectively as shown in Table 4.3.

Table 4.3: Data split

S/No	Split	Training		Split	Testing	
	%	Positive	Negative	%	Positive	Negative
1	50	2137	792	50	2136	791
2	60	2564	950	40	1709	633
3	70	2991	1108	30	1282	475
4	80	3418	1266	20	855	317
5	90	3846	1425	10	427	158

4.3 COVID-19

4.3.1 Dataset

A) COVID-19 Vs Healthy CXR images

Currently, there are closed to 80 million confirmed cases of patients suffering or suffered from COVID-19 disease. However, the number of CXR images available online on public database are limited. We acquired COVID-19 CXR images from 2 database. (1) 153 radiographs from GitHub website (available at <https://github.com/ieee8023/covid-chestxray-dataset>) and (2) 219 CXR images from Kaggle website (also available at <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>). Combining the 2 datasets to make up 371 total COVID-19 CXR images plus 1341 healthy CXR images from the same website.

B) Multiclass

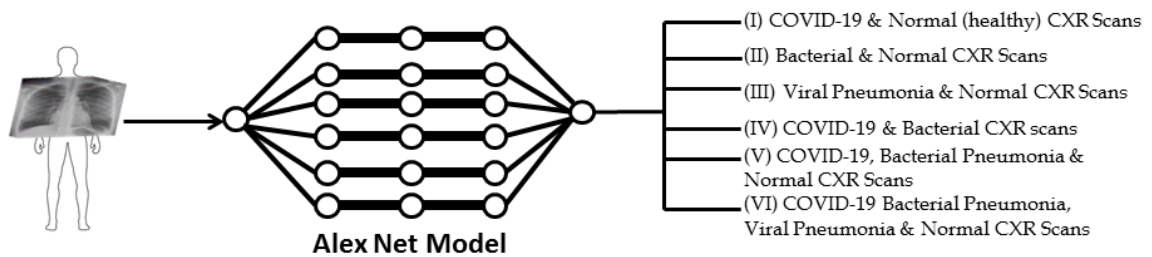


Figure 4.4: Workflow for multiclass classification

The dataset for COVID-19 and part of healthy CXR images are stated in COVID-19 Vs Healthy CXR images as well as other viral strains-causing pneumonia stated in pneumonia dataset acquired from Kermany et al., 2018. Moreover, 4274 bacterial pneumonia radiographs were obtained from Kaggle website (available at

<https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>). The workflow of the model is presented in Figure 4.4. Figure 4.5 shows the classes of pneumonia (positive) and negative radiographs while the number of each of the dataset is presented in Table 4.4.

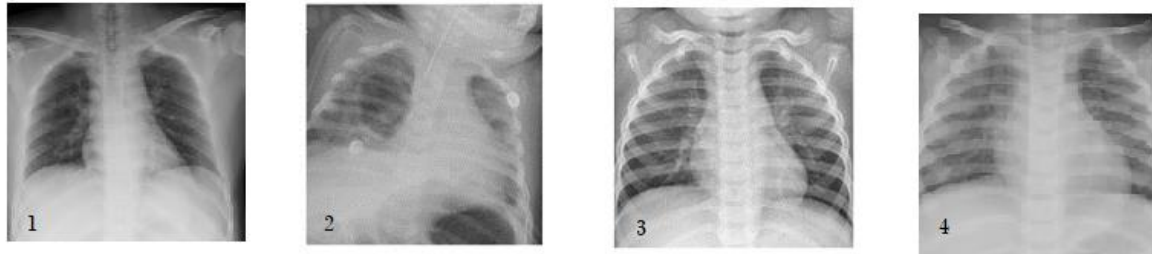


Figure 4.5: Dataset description: (1) COVID-19 (2) non-COVID 19 viral pneumonia (3) Healthy CXR scan (4) bacterial pneumonia

Table 4.4: Dataset Description for all classes

Image type	Total number of radiographs
Positive of COVID-19	371
Positive of other viral strains	4237
Positive with bacterial strains	4078
Normal	2882

4.3.2 Data split

The datasets are categorized based on binary classification and multiclass classification. Each dataset is partitioned into 70% for training the model and 30% for evaluation of the model on unseen dataset as shown in Table 4.5.

Table 4.5: Dataset split for multiclass

Model	Training 70%		Validation (30%)	
Non-COVID-19 VP and Healthy BP and Healthy	Positive	Negative	Positive	Negative
	2966	2017	1271	965
	Positive	Negative	Positive	Negative
	2853	2017	1225	965

COVID-19 and Healthy	COVID-19		Healthy		COVID-19		Healthy	
	260		2017		111		965	
COVID-19 and Non-COVID-19 VP	Positive		Non-COVID-19 VP		Positive		Positive	
	260		2966		111		1271	
COVID-19, BP and Healthy	C-19	BP	Negative		COVID-19	BP	Negative	
	260	2853	2017		111	1225	965	
COVID-19, Non-COVID-19 VP, BP and Healthy	C-19	BP	Non-C-19 VP	Negative	C-19	BP	Non-C-19 VP	Negative
	260	2853	2966	2017	111	1225	1271	965

4.4 Model Training

For both pneumonia and tuberculosis model training using 70% of dataset, we utilized a program known as “Matlab” which is run on a PC with the following descriptions: an intel® Core i7-3537U, window-64-bit, GPU and 8GB random access memory (RAM). The remaining 30% of the dataset split was reserved for evaluation of the model on unseen dataset (aka a testing dataset). Pretrained AlexNet model is utilized in this experiment because of its high precision in extracting feature and image discrimination. Figure 4 shows the AlexNet architecture used to classify both pneumonia Xray images and tuberculosis microscopic slide. AlexNet model architecture is based on different layers such as convolutional layers (which contain 5 convolution (CONV)) blocks made up of CONV-filters size 3x3 without undergoing padding operation and 2x2 window size for maximum pooling process. The final 3 layers are classified into 2; 2 FCL and lastly, an output layer

(OL). Other concepts regarding AlexNet are FM and BN. For classifying output into different categories, SoftMax activation function is employed in place of other classifiers such as SVM. In terms of gradient descent, a minibatch optimization is employed in order to optimize the model. Moreover, the training of the model is reached 20 epochs and learning rate of 0.0001.

The deep learning architecture utilized in this study is a CNN model known as AlexNet, a form of supervised machine learning which is the most popular ML techniques in which data are labelled and the model or network learn features to identify patterns in data for prediction or classification.

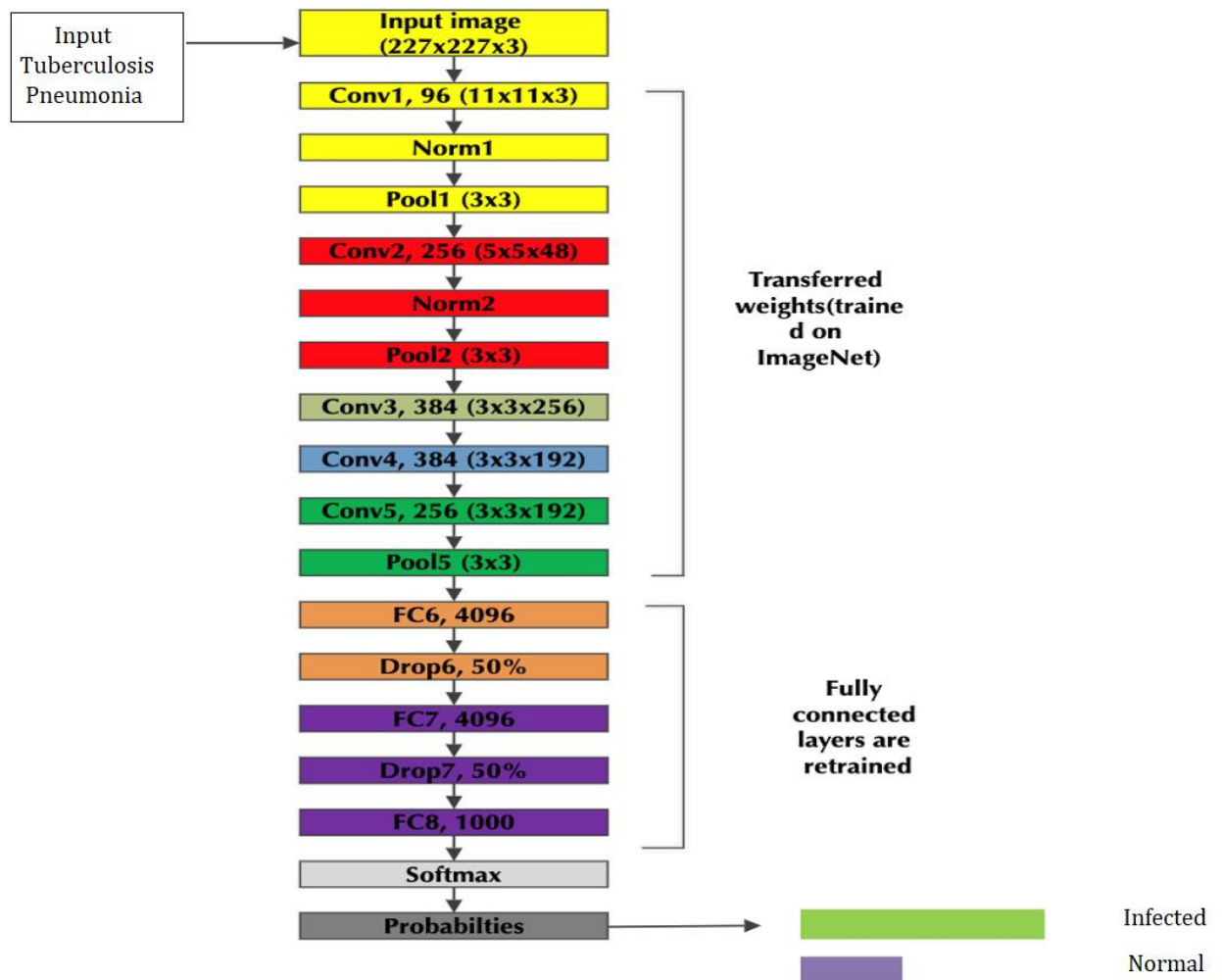


Figure 4.8: Classification of Medical Images using Pretrained AlexNet Model

Training ML models required series of steps depending on the deepness of the model. I.e., the deeper the model the more computation and number of layers. In this study, a Pretrained

AlexNet model is utilized which has 8 Layers as shown in Figure 4.6. The series of steps involved for training AlexNet model involves the following:

Step 1: Inputting: This is the first conducted in training AlexNet Model where images are fed into the model by creating image pathway using the code line below;

```
imds = imageDatastore('training', ...  
    'IncludeSubfolders',true, ...  
    'LabelSource','foldernames');
```

Step 2: Labelling: Once the model located the paths of the image folders, a 2 lines codes is used to label the data into training and validation and the num groups, i.e., for binary classification, num grouping result in a response equal 2 (e.g., positive and negative), while in the case of 4-way classification of pneumonia classes and normal CXR images, it results into num number equal to 4.

```
>> [imdsTrain,imdsValidation] = splitEachLabel(imds,0.7,'randomized');  
>> numTrainImages = numel(imdsTrain.Labels);
```

Step 3: Model Specification and Layers: For Matlab to run a model, 2 lines codes are used to specify the model name by using net equal to the model (AlexNet, SqueezeNet, MobileNet, GoogleNet etc.), these models have different layers and operations such as convolution, pooling, normalization etc. By adding the second code, Matlab will display all the network's layers, operations, filter size etc.

```
>> net = alexnet;  
>> net.Layers
```

Step 4: Input Size: As mentioned before, each network has a specified input size, for AlexNet, image with input size 227X227x3 (which means both vertical and horizontal pixels dimensions are 227) and 3 specify the color channel known as RGB (Red, Green and Blue). While for GoogleNet, image with input size 224X224X3 is utilized.

```
>>inputSize = net.Layers(1).InputSize
```

Step 5: Transfer Learning: Since the model is a pretrained model, several line codes are used:

```
>> layersTransfer = net.Layers(1:end-3);  
>> numClasses = numel(categories(imdsTrain.Labels))  
>> layers = [  
    layersTransfer
```

```

fullyConnectedLayer(numClasses,'WeightLearnRateFactor',20,'BiasLearnRateFactor',20)
    softmaxLayer
    classificationLayer];
>> pixelRange = [-30 30];
imageAugmenter = imageDataAugmenter( ...
    'RandXReflection',true, ...
    'RandXTranslation',pixelRange, ...
    'RandYTranslation',pixelRange);
augimdsTrain = augmentedImageDatastore(inputSize(1:2),imdsTrain, ...
    'DataAugmentation',imageAugmenter);
>> augimdsValidation = augmentedImageDatastore(inputSize(1:2),imdsValidation);

```

Step 6: Training: To initiate training, learning rate, epoch size and iterations are set before kick starting the training using graph plot which shows number of iterations per epoch of training progress using 2 lines codes:

```

>> options = trainingOptions('sgdm', ...
    'MiniBatchSize',10, ...
    'MaxEpochs',20, ...
    'InitialLearnRate',1e-4, ...
    'ValidationData',augimdsValidation, ...
    'ValidationFrequency',3, ...
    'ValidationPatience',Inf, ...
    'Verbose',true, ...
    'Plots','training-progress');
>> netTransfer = trainNetwork(augimdsTrain,layers,options);

```

Testing and Validation:

Step 1: Loading: To test the performance of the model, testing dataset are loaded by creating image pathway using the code line below:

```

>> imdsTest = imageDatastore('Testing', ...
    'IncludeSubfolders',true, ...
    'LabelSource','foldernames');

```

Step 2: Testing: After the images are loaded, a 3 lines codes are used to test the model based on accuracy:

```
>> [YPred,scores] = classify(netTransfer,imdsTest);  
>> YValidation = imdsTest.Labels;  
>> accuracy = mean(YPred == YValidation)
```

Step 3: Sensitivity and Specificity: In order to check the positive and negative rate classification of the model, a 3 lines codes are inputted:

```
>> y = grp2idx(YValidation);  
>> test = grp2idx(YPred);  
>> classperf(y,test)
```

4.5 Cross Validation

Cross validation (CV) is a vital technique utilized in ML for parameter selection and evaluation of learning performance and prediction. Based on this experiment, we utilized K-fold Cross validation approach where the datasets are split into K sets of equal size (i.e., K = 10). In each K sets K-1 is used as training dataset and 1 set is used as validation dataset. Training of the dataset is repeated for K number of times (i.e., n = k) (Fan and Hauser 2018). The average performance of the training and testing dataset is computed as the evaluation index for the models. This approach is very efficient especially when there are limited number of samples as it takes advantage of the whole dataset (Men et al., 2018). Hence, Cross validation dataset for both pneumonia and tuberculosis are presented in the supplementary file

4.6 Evaluation and Confusion Matrix

To evaluate the performance of the learned models, some certain parameters are utilized; accuracy, Precision also known as sensitivity and specificity or recall. Accuracy is defined as the ratio of properly classified images over total sum of images; it is also described as the sum of precision and recall. For evaluating the accuracy and loss of the learned model the resulting formulas are employed:

$$Loss = -\frac{1}{n} \sum_{i=1}^n \log PC \quad (2.1)$$

$$Accuracy = \frac{C}{N} \quad (2.2)$$

Where N is the overall number of images during training and testing, n is the number of images and PC is the probability of the correctly classified images.

Confusion matrix is the common approach used for evaluation of model performance based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TPs is the number of samples that are correctly identified by the model as positive cases or number of cases who actually have pneumonia or tuberculosis according to each model. TNs is the number of samples that are correctly identified by the model as negative cases or number of cases who are actually healthy (normal) and classified as negative according to each model. FPs are the number of samples that are incorrectly classified as negative by the model or number of cases that are actually negative (normal or healthy) but classified as pneumonia or tuberculosis according to each model. FNs are the number of samples that are incorrectly classified as positive by the model or number of cases that are actually positive (pneumonia or tuberculosis) but classified as normal or healthy according to each model as shown in Table 4.6.

True Positive rate (Sensitivity) is the portion of positive cases or samples which are precisely classified as positive sample (i.e., it describes the ration of positive cases that are correctly identified as positives).

$$Sensitivity = \frac{TPs}{TPs+FNs} \quad (2.3)$$

False positive rate (FPR) also known as Specificity is the portion of positive cases or samples which are wrongly classified as positive samples (i.e., it describes the ratio of negative samples that are incorrectly classified as positives).

$$Specificity = \frac{TNs}{TNs+FPs} \quad (2.4)$$

Table 4.6: Confusion matrix

Predictions	Actual Positive	Actual Negatives
Positive Predictions	TP	FP
Negative Predictions	FN	TN

CHAPTER 5

PERFORMANCE EVALUATION

The models are trained using pretrained AlexNet which comprise of 8 total layers, 5 CONV layers and 3 FCLs. The models were able to learn various features from all the convolutional layers. Each convolutional layer performs separate form of feature selection (point, edges, structure etc.). The neurons in convolution 1 are very active in spotting features while the final convolutional layer (Conv 5) is very active in level detection and descriptive features. However, all the convolutional layers learn features differently which enhance the model learning efficiency. The last layer (fully connected layer 8) classifies the images into binary (in tuberculosis, Mycobacterium tuberculosis vs healthy microscopic slide, COVID-19 vs Normal CXR images, other viral strains-causing pneumonia vs normal radiographic images) and into 3 and 4 classes (in classification of different types of pneumonia and healthy radiographs). Some of the parameters for training the models include:

1. Iterations: which is define as the total number of batches required to complete a single epoch (i.e., circle).
2. Epoch: is termed as the number of iterations multiply by the size of the batch which is further divided by total number of images used in training.
3. Iterations per epoch is calculated based on dividing maximum iterations by epoch.

5.1 Tuberculosis

5.1.1 Performance Evaluation

Training of the model resulted in maximum iterations of 3440, 20 completed epochs and 172 iterations per epoch. 0.0001 learning rate was achieved and 0.0127 error rate. In terms of performance evaluation, the model achieved training accuracy of 99.19%, validation accuracy of 98.73%, sensitivity of 98.59% and specificity of 98.48%. The description of the result is presented in figure 5.1, learning curve in Figure 5.2 and Table 5.1.

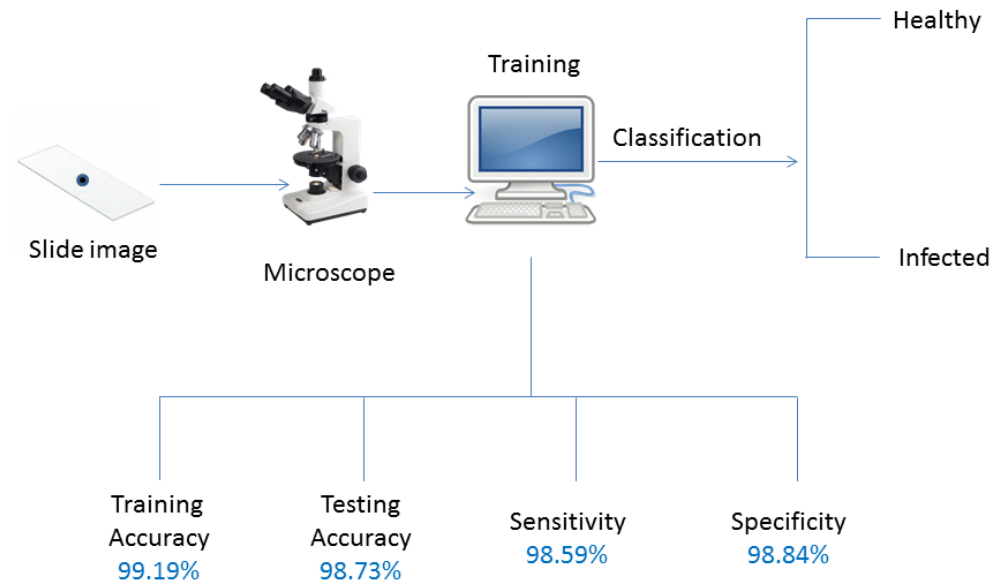


Figure 5.1: Classification of tuberculosis and healthy microscopic slides using pretrained AlexNet Model

Table 5.1: Model Learning Parameters and Performance

Learning Parameters	Values
Training ratio (%)	70
Learning rates	0.0001
No of epochs	20
Training accuracy (%)	99.19
Testing accuracy (%)	98.73
Sensitivity	98.59
Specificity	98.84
Achieved mean square error	0.0127

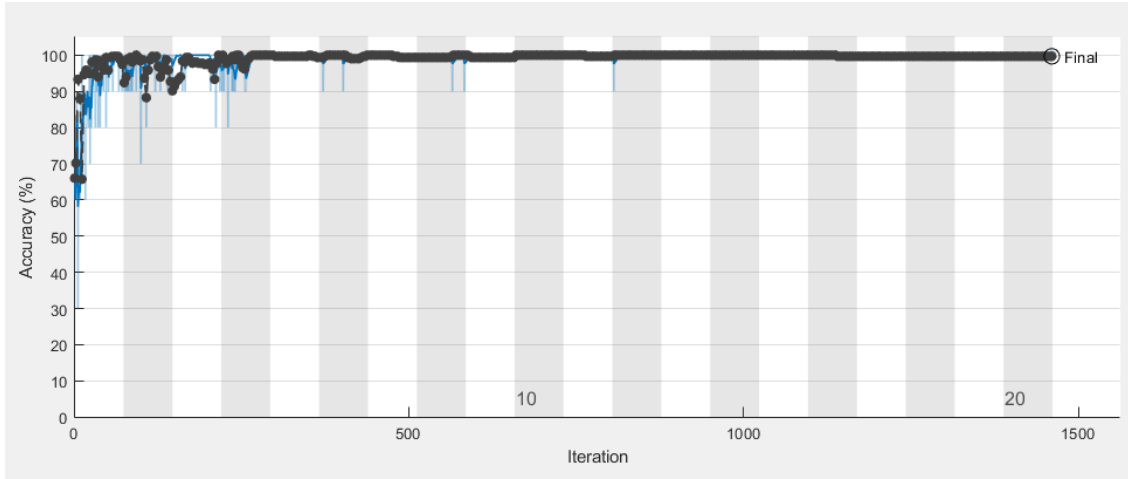


Figure 5.2: Learning Curve of AlexNet network

5.1.2 Cross Validation

CV was carried out to assess the overall performance of the model by utilizing every K-Fold as training and testing set. The model achieved average performance of 99.28% training accuracy and 98.29% testing accuracy. Based on sensitivity and specificity, the model achieved 99.20% and 99.39% respectively. Comparing the result of general dataset with CV, the average performance of CV ranked higher in terms of training accuracy, testing accuracy and sensitivity while general dataset performed better in terms of specificity. The means (average) result obtained from all the folds are presented in Table 5.2.

Table 5.2: Cross validation result

K Fold	Training Accuracy	validation	Testing Accuracy	Sensitivity	Specificity
1	98.15	0.9815	98.09	0.9677	1.000
2	99.07	0.9907	97.45	0.9852	0.9963
3	99.81	0.9981	97.45	1.0000	0.9963
4	99.07	0.9907	98.73	1.0000	0.9815
5	98.89	0.9889	98.09	0.9889	0.9889
6	100.00	1.0000	100.00	1.0000	1.0000
7	99.44	0.9944	98.64	1.0000	0.9868
8	99.44	0.9944	98.73	0.9963	0.9926
9	99.63	0.9963	97.13	0.9963	0.9963

10	99.26	0.9926	98.54	0.9859	1.0000
Average	992.76/10	9.99276/10	982.85/10	9.92303/10	9.9387/10
	99.28	0.9928	98.29	0.9920	0.9939

5.1.3 Machine Vs Human

Comparing machine vs human is conducted in order to test the network the performance. 30 unseen images are given to the model, beginners and certified pathologist. The model was able to classify the images correctly outperforming both pathologists. The result of the performance is shown in table 5.3.

Table 5.3: Performance evaluation of machine vs humans

N = 30			
Participants	Sensitivity (%)	Specificity (%)	Accuracy (%)
Machine (Model in Ex C)	100	100	100
Beginner 1	100	85.71	92.86
Beginner 2	94.74	92.31	96.15
Certified 1	94.74	100	97.37
Certified 2	100	92.31	96.15

5.1.4 Comparison with state of art

We obtained a testing accuracy of 97.64% using general dataset and testing accuracy of 98.29% using the average accuracies of Cross validation result. Our model has achieved a better accuracy than the study conducted by Xiong et al 2018 who built a model from scratch and trained using limited number of cases. The model achieved 83.65% specificity and 97.94% sensitivity compare to our model which achieved 98.67% sensitivity and 100% specificity using general dataset and 99.20% sensitivity and 99.39% specificity using average performance of cross validation. The recent study carried out by Muyama et al., 2020 achieved 86.7% accuracy which ranked lower in comparison with our model. Moreover, the experiment conducted by Panicker et al 2018 using microscopic sputum smear images trained on a CNN model achieved sensitivity of 97.13% and specificity of 78.4%. Table 5.4 depict the comparison between model outcome and related research.

Table 5.4: Comparison between model outcome and state-of-the-art

Reference	Model	Accuracy (%)	Sensitivity (%)	Specificity
Khan et al 2019	CNN	94.0	-	
Xiong et al 2018	CNN	97.0	97.94	83.65
Smith et al 2018	CNN	95.0	-	
Costa et al 2015	SVM+CNN	96.8	-	
Panicker et al 2018	CNN	-	97.13	78.4
El-Melegy et al 2019	FRCNN	98.3	82.6	
	FRCNN+CNN	98.4	85.1	
Muyama et al., 2020	Pretrained (ResNet, GoogleNet and VGGNet)	86.7	-	-
Our Result	Pre-trained AlexNet	98.73	98.59	99.39
CV	Pre-trained AlexNet	99.29	98.20	99.39

5.2 Pneumonia

5.2.1 General Dataset

The model was trained using different set of data split. The entire dataset (i.e., 5856 total chest Xray images) were partition into 50:50, 60:40, 70:30, 80:20 and 90:10 percentages for training and testing. Matlab is employed to trained the dataset with 0.0001 learning rates, 20 epochs and 5740 number of iterations. The model performances were evaluated based on accuracy, sensitivity and specificity.

Based on 50:50 data split, the model achieved 97.98% training accuracy, 97.94% testing accuracy, 96.21% sensitivity and 99.00% specificity. As expected, by increasing number of training dataset to 60% and reducing number of testing dataset to 40%, the model achieved 98.94% training accuracy, 98.95% testing accuracy, 99.09% sensitivity and 98.81% specificity. Considering the fact that training with large number of datasets mostly result in higher training accuracy. However, scientist recommend the use of 70:30 or 80:20 data split in order to allow the model learn efficiently and performed accurately on testing datasets (Zhang et al., 2018; Shorten & Khoshgoftaar, 2019).

Based on 70:30 data split, the model achieved 99.19% training accuracy, 98.73% testing accuracy, 98.59% sensitivity and 98.84% specificity. Training the model using 80% of the entire dataset and testing using 20% result in 99.36% training accuracy, 100% testing accuracy, 99.11% sensitivity and 99.66% specificity. While training using 90% of the entire dataset and testing using 10% result in 99.86% training accuracy, 100% testing accuracy, 99.11% sensitivity and 100% specificity. The performance evaluation of models trained based on different data split is presented in table 5.5 and Figure 5.3.

Table 5.5: Performance evaluation of models trained using different data split

Split	Training Accuracy	Testing Accuracy	Sv	Sf
50-50	97.96	97.94	96.71	99.00
60-40	98.94	98.95	99.09	98.81
70-30	99.19	98.73	98.59	98.84
80-20	99.36	100.00	99.11	99.66
90-10	99.86	100.00	99.70	100.00

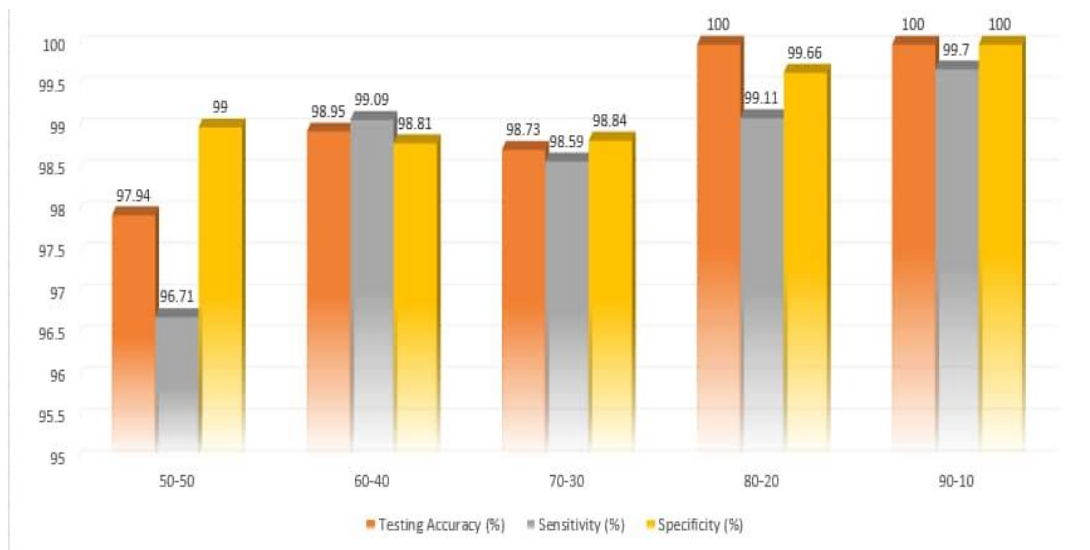


Figure 5.3: Performance evaluation of AlexNet models trained using different data split

So many studies have shown that using higher number of dataset results in higher model performance (i.e., increasing training dataset led to increase in training accuracy). As shown

in Table 5.4 and Figure 5.3, training accuracy increase from 97.96% in 50:50 data split to 99.86% in 90:10 data split. The same applied to testing accuracy. As a result of training the model using large amount of dataset, it learns to classified both positive and negative cases correctly, resulting in higher testing accuracy (which increase from 97.94% in 50:50 data split to 100% in 90:10 data split). A similar approach was adopted by Prashanth et al., 2020 who partitioned dataset into different data split (from 50:50 to 90:10) for model-training and validation respectively. The study reported higher accuracy for data split above 60:40.

5.2.2 Cross Validation

We divided the dataset into 10 folds, in each fold we used K-1 as validation dataset and the remaining as training dataset (i.e., 9 folds), separate untested new images are used as testing dataset to ascertain the performance of the TL architecture. The accuracy of training dataset is compared with the testing dataset to evaluate performance and overfitting. Sensitivity and specificity are other parameters used to check false positive rate and true positive rate likelihood.

Comparing all the folds in terms of performances, testing accuracies of all the folds are lower than training accuracy as expected except for K-4. This is true for majority models trained using deep learning as reported by Keskar et al.,2016. Comparing the means of both training and testing accuracies, testing accuracy ranked lower (i.e., 96.04%) than training accuracy (i.e., 97.07%). However, there are so many variations in terms of sensitivity and specificity in the 10-folds. Moreover, based on the result achieved in terms of positive and negative rate, it shows that the models learned how to discriminate between pneumonia and healthy chest Xray images. The result of cross validation is presented in Table 5.6.

Table 5.6: Cross validation result for Pneumonia

K Fold	Tr(A)*	V*	Ts(A)*	Sv*	Si*
1	98.35	0.9835	96.67	0.9800	0.9846
2	96.78	0.9678	94.71	0.9767	0.9650
3	97.72	0.9772	96.55	0.9867	0.9743
4	97.56	0.9756	94.71	0.9567	0.9815
5	97.72	0.9772	98.16	0.9567	0.9835
6	97.48	0.9748	94.14	0.9867	0.9712

7	96.86	0.9686	93.45	0.9800	0.9650
8	98.35	0.9835	96.21	0.9633	0.9897
9	98.27	0.9827	95.63	0.9867	0.9815
10	97.88	0.9788	97.13	0.9633	0.9835
Average	976.97/10	9.76970/10	960.36/10	9.37368/10	9.7798/10
	97.70	0.9770	96.04	0.9734	0.9779

5.2.3 Comparison between General Data Split and Cross Validation

In general dataset, the whole images were split into 50:50, 60:40, 70:30, 80:20 and 90:10 for model-training and validation respectively resulting in irregular training accuracies, testing accuracies, sensitivities and specificities. However, based on CV, the average performance of the model result in 97.70% training accuracy, 96.04% testing accuracy, 97.34% sensitivity and 97.79% specificity. This shows that the results achieved in cross validation ranked lower than all the splits as presented in Table 5.7.

Table 5.7: Performance evaluation of models trained using different data split

Split	Training Accuracy	Testing Accuracy	Sv	Sf
50-50	97.96	97.94	96.71	99.00
60-40	98.94	98.95	99.09	98.81
70-30	99.19	98.73	98.59	98.84
80-20	99.36	100.00	99.11	99.66
90-10	99.86	100.00	99.70	100.00
CV	97.70	96.34	97.34	97.79

Radiologist have been relying radiological images for interpreting pneumonia based on the presence of infiltrates (white spots in the patient's lungs) to identify or interpret the presence of the infection and other complications such as pleural effusions or abscesses. On the other hand, pathologist rely on staining techniques for visualization of tuberculosis bacilli in sputum smear slides. For both radiologist and pathologist, these approaches can be very tedious for large images and thus can lead to misinterpretation. The use of computer aided

diagnosis (CAD) which was introduced in 1990s offer a simple, reliable, precise and fast approach of interpreting results related to medical images. CAD approach assist pathologist and radiologist in identifying disease and healthy images while preventing misinterpretation.

5.2.4 Comparison with State of Art

The use of CNN to classify and characterize X-ray images has shown a better accuracy and precision than some radiologist. Since the development of deep neural network, scientist have been utilizing different CNN architectures such as ResNet, Inceptions, AlexNet, GoogleNet, SqueezeNet, VGGNet 16 and 17 etc. and other networks built from scratch to detect diseases such as pneumonia in x-ray images and tuberculosis in acid-fast stain images. These computer models are developed based on mathematical algorithms to solve problems such as predictions and image classification using probability score (Matsugu et al 2003). Comparing our result with state of art, for pneumonia we obtained a testing accuracy of 96.04% using general dataset and testing accuracy of 97.70% using the average accuracies of Cross validation result. Our model has achieved a better accuracy than the study conducted by Stephen et al 2019 using the same dataset but different model that is built from scratch which achieved average accuracy of 94.81%. Saraiva et al 2019 utilized the same dataset with our study, the authors split the dataset into 5 K-folds and achieved 95.30% average accuracy while we split our dataset into 10 k-folds and achieved average accuracy of 97.70%. Rajaraman et al 2018 utilized VGG-16 to classify both bacterial and viral pneumonia. The models achieved 96.2% and 93.6% compare to our model that achieved 96.43 for general dataset using AlexNet models. However, studies that utilized large amount of dataset of over 100 thousand have achieved a higher accuracy than our study such as Zech et al 2018 who utilized Deep NN to trained 158,323 x-ray images and Rajpurkar et al who developed a 121 CNN called CHeXNet and trained the model using more than 100 thousand frontal view X-ray images. Table 5.8 depict the comparison between model outcome and related research.

Table 5.8: Comparison between model outcome and state-of-the-art

Reference	No of Dataset	Model	A/AUC	Sv	Sf
-----------	---------------	-------	-------	----	----

70:30	5856	PA	98.73	98.59	98.84
CV	5856	PA	97.35	97.35	97.78
Stephen et al. 2019	5856	CNN	94.81	-	-
Chouhan et al. 2020	5856	PA	92.86	-	-
Saravia et al. 2019	5856	CNN	95.30	-	-
Rajaraman et al. 2018	5856	CNN	92.2, 93.6	-	-
Kanaparth i et al. 2019	108,948	PA	0.6333	-	-
Rajpurkar et al. 2017	100,000	CheXNet	0.8887	-	-

5.3 COVID-19 and Other Pneumonia

5.3.1 Binary Classification

The classification of COVID-19 vs normal radiographs, COVID-19 vs other viral strain-causing pneumonia, other viral strain-causing pneumonia vs normal CXR images, bacterial pneumonia vs healthy CXR. Table 5.9 and Figure 5.4 present the classification.

For all the classifications, datasets are trained based on 70% and tested using 30%. For classifications of COVID-19 and normal CXR images, the performance of the model resulted in 99.71% training accuracy, 99.16% testing accuracy-score, 97.44% precision and 100% recall. For 2-way classification of COVID-19 and other viral strain-causing pneumonia, the performance of the model resulted in 99.57% training accuracy score, 99.62% testing accuracy score, 90.63% precision and 99.89% recall. Based on 2-way discrimination of other viral strain-causing pneumonia and normal radiographs, the performance of the model resulted in 96.43% training accuracy score, 94.05% testing

accuracy score, 98.19% precision and 95.78% recall. For classification of bacterial pneumonia and normal CXR images, the performance of the model resulted in 95.28% training accuracy score, 91.96% testing accuracy score, 91.94% precision and 100% recall.

5.3.2 Multiclass

A pilot study was initially conducted using equal number of datasets (i.e., 371 CXR images) for each class (COVID-19, other viral strain-causing pneumonia, bacterial pneumonia and normal radiographs). The justification for this amount is due to the fact that we acquired only 371 CXR images of COVID-19 pneumonia. However, the model performed poorly in terms of training accuracy score, testing accuracy score, precision and recall.

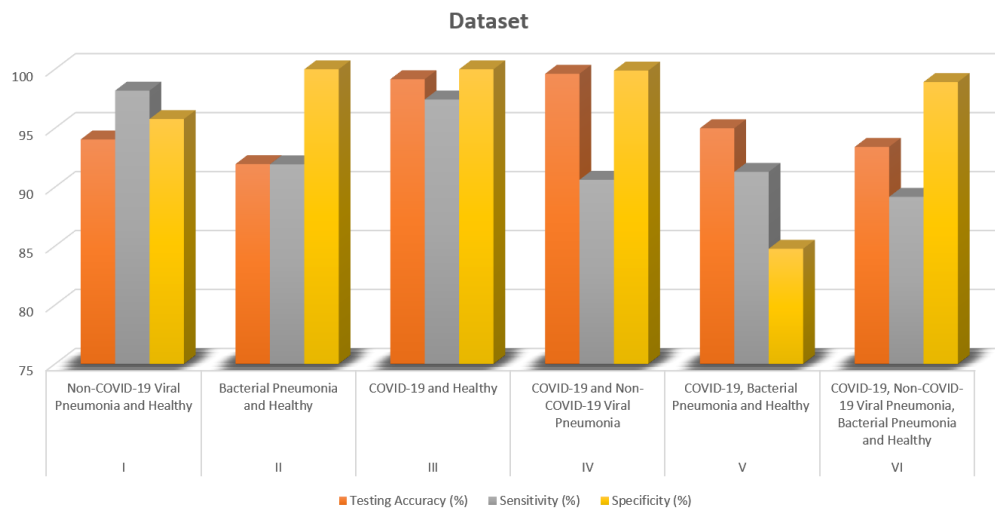


Figure 5.4: Performance Evaluation of models based on accuracy score, precision and recall.

The model was then trained using whole dataset partitioned into 70% for training and 30% for testing for each class as shown in Table 4.4. The model was trained according to 3-way classification (COVID-19, bacterial pneumonia and normal radiographs), the model performed very well in terms of training accuracy (97.40%), testing accuracy (95.00%), sensitivity (91.30%) and poorly in terms specificity (84.78%) compare to 2 classes. By adding non-COVID-19 viral pneumonia to the 3 classes above (i.e., making it 4 classes) the model achieved 94.18% training accuracy, 93.42% testing accuracy, 89.18% sensitivity and 98.92% specificity as presented in Table 5.9 and Figure 5.4.

Table 5.9: Performance evaluation of binary classes and multiclass

Dataset	Training Accuracy (%)	Testing Accuracy (%)	Sv (%)	Sf (%)
other viral strain- causing pneumonia and Normal	96.43	94.05	98.19	95.78
Bacterial pneumonia and Normal	95.28	91.96	91.94	100.00
COVID-19 and healthy	99.71	99.16	97.44	100.00
COVID-19 other viral strain-causing pneumonia	99.57	99.62	90.63	99.89
COVID-19, bacterial pneumonia and normal	97.40	95.00	91.30	84.78
COVID-19, other viral strain-causing pneumonia, bacterial pneumonia and normal	94.18	93.42	89.18	98.92

5.3.3 Comparison with State of Art

Several studies have attempted to distinguished COVID-19 pneumonia from normal radiographs. However, few studies discriminate between COVID-19 pneumonia and other types of pneumonia from CXR scan images. Li et al., 2020 grouped both viral pneumonia and bacterial pneumonia as Community Acquired Pneumonia (CAP). Though, we deemed this classification unnecessary in terms of AI-models due to the facts that bacterial and viral pneumonia differs and majority of viral pneumonia have the same appearance in CXR scans. Based on binary classification of COVID-19 pneumonia from normal radiographs, our model achieved higher performance compared with studies that employed lesser amount of dataset such as Narin et al., 2020, Mahmud et al., 2020, Abbas et al., 2020 and Wang S et al., 2020. Among the studies we reviewed, Abbas et al., 2020 achieved a higher sensitivity

(97.91%) compare to our model. This study achieved higher sensitivity due to the use of TL on DL with more of layers. Also, Mohammadi et al., 2020 achieved higher performance in terms of sensitivity (98%) compare to our model. The scores are attributed to large amount of dataset resulted from data augmentation and the use of deep learning models with higher number of models than AlexNet. Different studies confirmed that TL models perform efficiently on less amount of data compare with models developed from scratch which required large amount of training datasets (Shin et al., 2016; Ravishankar et al., 2016; Yu et al., 2017).

The discrimination of COVID-19 from other viral strain-causing pneumonia is critical in order to avoid miss-diagnosis. Differentiating between COVID-19 pneumonia and other viral strain-causing pneumonia is merely impossible for Radiologist due to the fact that both are viral pneumonia. It is quite easier for AI-driven models to discriminate between bacterial pneumonia and viral pneumonia (with COVID-19 pneumonia inclusive). This claimed is supported by study carried out by Chowdhury et al., 2020 who have shown that “models perform more efficiently on classification of COVID-19 pneumonia and normal radiographs compare to COVID-19 pneumonia and other viral strain-causing pneumonia.

The result achieved by our model is in line with this study. In terms of classifying COVID-19 pneumonia and other viral strain-causing pneumonia, our model achieved 90.63% sensitivity and 99.89% specificity compare with classification of COVID-19 pneumonia and normal radiographs where the model achieved 97.44% precision and 100% recall. Moreover, Narin et al., 2020 and Bai et al., 2020 also stated similarities in terms of appearance between COVID-19 pneumonia and non-COVID-19 viral pneumonia based on clinical and physiological prospective.

Majority of pneumonia occur as a result of invading viruses and pathogenic bacteria. In terms of viral pneumonia, majority of cases are caused by Influenza virus which is prevalence in pediatric patients. Differentiating between Influenza-causing pneumonia from healthy CXR images is crucial for proper diagnosis. Both CXR images of Influenza-causing pneumonia and healthy ones are made available by Kermany et al., 2018. Based on this dataset, our model performance resulted in 94.43% accuracy score, 98.19% precision and 95.78% recall compare to Stephen et al., 2019 with accuracy of 93.01%, Wang SS et al., 2020 with 73.1% accuracy, 74% sensitivity and 67% specificity, Xu et al., 2020 with 86.7%

accuracy and Rajaraman with 93.6% accuracy. However, only Saravia et al., 2019 achieved higher accuracy (95.30%) compare to our model. Table 5.10 shows the comparison between our model with previous studies.

Table 5.10: Comparison between our Result (BC) with the State-of-the-Art

Class	Reference	Result		
		Ac	S	Sf
COVID-19 and other viral strain-causing pneumonia	Mahmud et al., 2020	96.9%	-	-
	Our model (COVID-19 and other VP)	99.62%	90.63%	99.89%
	Stephen et al., 2019	93.01%	-	-
	Wang SS et al., 2020	73.1%,	74%	67%
	Saravia et al., 2019	95.30%	-	-
	Xu et al., 2020	86.7%.	-	-
pneumonia and normal datasets	Rajaraman et al., 2018	96.2% for BP and 93.6% for Non-COVID-19 VP	-	-
	Our model (other viral strain-causing pneumonia and normal datasets)	94.43%	98.19%	95.78%
	Narin et al., 2020	97% for InceptionV3 and 87% for Inception-ResNetV2	-	-
	Mahmud et al., 2020	97.4%	-	-
COVID-19 and normal datasets	Abbas et al., 2020	95.12%	97.91%	91.87%
	Wang et al., 2020	73.1%	74.00%	67.00%

Mohammedi et al., 2020	99.1%	98.0%	100%
Our model (COVID-19 and normal datasets)	99.16%	97.44%	100%

Based on multiclass classification (i.e., 3-way and 4-way classification), our model performed better in terms of accuracy (94%), sensitivity (91.30%), specificity (84.78%) for 3-way classification compare to result obtained by Mahmud et al., 2020 with 90% accuracy, Ehtioui et al., 2020 with 91.3% accuracy and Ozturk et al., 2020 with 87.02% accuracy. However, we couldn't compare our result with the study reported by Li et al., 2020 due to the fact that they classified each class separately with 90% precision and 96% recall for COVID-19 and 87% precision and 92% recall for CAP. The study reported by Quan et al., 2020 achieved higher performance compare to our work, this is due to the training of the model with large amount of dataset (i.e., 14 thousand) and the fusion of models which have shown to perform better than single model. Currently, there are no articles that addressed 4-way classification. Multiclass comparison between our model and the current state-of-the-art is presented in Table 5.11.

Table 5.11: Multiclass Comparison between our Result with the State-of-the-Art

Reference	Dataset	Result		
		Ac	Sv	Sf
Li et al., 2020	4352 CT scans (1292 of COVID-19, 1735 of CAP and 1325 normal CT scans)	-	90% for COVID-19 87% for CAP	96% for COVID-19 92% for CAP
Mahmud et al., 2020	1493 non-COVID-19 VP, 305 COVID-19 P, 2780 BP	90%	-	-
Ehtioui et al., 2020	500 non-COVID-19 VP, 500 COVID-19 P and 500 normal CXR images	91.34%	-	-

Quan et al., 2020	14 thousand CXR images (COVID-19, bacterial pneumonia, viral pneumonia and normal)	99.32%	98.55	100%
Our model (3- way classification)	371 COVID-19, 4078 BP and 2882 healthy	94.00%	91.30%	84.78%
Our model (4- way classification)	371 COVID-19, 4237 non- COVID-19 VP, 4078 BP and 2882 healthy	93.42%	89.18%	98.92%

5.3.4 Comparison between CRISPR-based Biosensors and CAD

CRISPR-based biosensors have shown to be the most sensitive and specific method for detection of infectious disease such as Tuberculosis, COVID-19 pneumonia, bacterial pneumonia and other viral pneumonia. However, this technique requires sophisticated instrument mount in laboratories, chemical reagents, highly trained personnel and longer processing and detection time. CAD on the other hand rely on PC or desktop and experts for training and running of the model and high amount of training dataset to reach optimum accuracy. This approach can be used as a confirmatory approach or as substitute for molecular diagnosis due to it low sensitivity and specificity compare to CRISPR-based method.

Table 5.12: Comparison between CRISPR-based Biosensors and CAD

Feature	CRISPR-based biosensor	CAD
Sample/Dataset	DNA/RNA isolated from patient's blood, urine, sputum or nasal swab	Microscopic slide images, CT scans and Chest X-ray images of patients
Amplification	PCR method, RPA	Data augmentation
Instruments/Analysis	Require sophisticated instrument mount in laboratories, chemical reagents and highly trained personnel	Require PC or desktop and experts for training and running of the model
Detection efficiency	Highly specific and sensitive	High accuracy for classification

Processing time	Diagnosis range between 1 hour to weeks depending on the pathogen and approach	Training takes time depending on the amounts of datasets, classification takes less than 1 minutes
Set-up Cost	Very expensive	Cheap

CHAPTER 6

CONCLUSION

Pneumonia and tuberculosis are among the common diseases that affects the lungs. Detection of these diseases are crucial due to their prevalence in underdeveloped countries and remote areas. These diseases are mostly diagnosed using bench diagnosis assays which utilize chemical reagent, trained pathologist and radiologist, longer procedure and heavy workload. The main goal of our study is to utilized DL approach to classify pneumonia based on CXR samples and tuberculosis based on microscopic slides images and to compare the efficiency between CAD approach and the use of CRISPR-based biosensor. In this framework, we employed pretrained AlexNet architecture as a form of TL technique.

Based on discrimination of tuberculosis and healthy microscopic slide based on the whole dataset (i.e., 7:3 ratio for learning and validation respectively), the model achieved training accuracy of 99.19%, testing accuracy-score of 98.73%, precision of 98.59% and recall of 98.48%. In terms of cross validation based on 10-K folds, the model achieved average performance of 99.28% training accuracy and 98.29% testing accuracy. To check the efficiency of the model, machine vs human is carried out. 30 unseen images are given to the model, beginners and certified pathologist. The model was able to classify the images correctly outperforming both pathologists.

For classification of pneumonia using x-ray images, models are trained based on different splits (50:50, 60:40, 70:30, 80:20 and 90:10) and cross validation based on 10-k folds to differentiate between viral pneumonia and healthy patients. Based on 50:50 data split, the model achieved 97.98% training accuracy, 97.94% testing accuracy, 96.21% sensitivity and 99.00% specificity. 60:40 splits, the model achieved 98.94% training accuracy, 98.95% testing accuracy, 99.09% sensitivity and 98.81% specificity. For 70:30 split, the model achieved 99.19% training accuracy, 98.73% testing accuracy, 98.59% sensitivity and 98.84% specificity. For 80:20 split, the model achieved 99.36% training accuracy, 100% testing accuracy, 99.11% sensitivity and 99.66% specificity. For 90:10 split, the model achieved 99.86% training accuracy, 100% testing accuracy, 99.11% sensitivity and 100%

specificity. Based on cross validation, the models achieved average performance of 97.70% training accuracy, 96.04% validation accuracy-score, 97.34% precision and 97.79% recall. The outbreak of infection caused by a strain of coronavirus has led to global health and economic crisis. Medical practitioners rely on the RT-PCR method which is clouded by so many challenges, such as high cost of the test, laborious, tedious and high rate of false positive results. To provide an alternative, scientist turned to X-ray images for detection of pneumonia caused by COVID-19 disease. In the light of these challenges, we trained our models to classified COVID-19 from healthy CXR images as well as other binary classification of pneumonia and multiclass (3-way classification and 4-way classification). For classification of disease causes by SARS-CoV-2 and other viral pneumonia and healthy cases, the model was able to discriminate between different classes based on binary classifications, multiclass (3-way and 4-way classifications) with high accuracies, sensitivities and specificities.

Our result is in line with the notion that CNN models can be used for classifying medical images with higher accuracy and precision. These models can now serve as a confirmation system for diagnosis of both pneumonia by and tuberculosis, maximizing miss diagnosis and offer an alternative to relieve the heavy and tedious workload experiencing by radiologist and pathologist in Near East University Hospital. Comparing CRISPR-based biosensors and CAD approaches, CRISPR-based biosensors have demonstrated to be among the most sensitive and specific method for diagnosis and screening of infectious disease. CAD on the other hand can be used as a confirmatory approach or as substitute for molecular diagnosis due to it low sensitivity and specificity compare to CRISPR-based method.

Among the challenges and limitations of our research include the lack of sufficient amount of dataset especially for radiographic images of COVID-19 pneumonia. For classification of pneumonia from normal CXR images, only frontal radiographs are employed without augmentation. Normally, frontal images are the types interpreted by radiologist without the need of rotation or color shift. Thus, with large amount of dataset we can utilize different pretrained architectures such as VGGNet, GoogleNet and ResNet. Moreover, the use of hybrid models or combining CNN models with other models (i.e., ensembled ones) or regression tools and classifiers can help improve performance classification of the models.

Another challenge encountered in this study is lack of laboratory study for detection of tuberculosis and pneumonia using CRISPR-based biosensors.

REFERENCES

- Abbas, A., & Abdelsamea, M. M. (2018, December). Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network. In 2018 13th International Conference on Computer Engineering and Systems (ICCES) (pp. 122-126). IEEE.
- Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2020). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. arXiv preprint arXiv:2003.13815.
- Abiyev, R. H., & Ma'aitah, M. K. S. (2018). Deep convolutional neural networks for chest diseases detection. *Journal of healthcare engineering*, 2018. <https://doi.org/10.1155/2018/4168538>.
- Abudayyeh OO, Gootenberg JS (2019) Chippingin on diagnostics. *TheCRISPR journal*100(102):69–71. <https://doi.org/10.1089/crispr.2019.29053.oma>
- Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in biology and medicine*, 100, 270-278. <https://doi.org/10.1016/j.combiomed.2017.09.017>.
- Aguiar, F. S., Torres, R. C., Pinto, J. V., Kritski, A. L., Seixas, J. M., & Mello, F. C. (2016). Development of two artificial neural network models to support the diagnosis of pulmonary tuberculosis in hospitalized patients in Rio de Janeiro, Brazil. *Medical & biological engineering & computing*, 54(11), 1751-1759. DOI 10.1007/s11517-016-1465-1.
- Ahsan, M., Gomes, R., & Denton, A. (2019, May). Application of a Convolutional Neural Network using transfer learning for tuberculosis detection. In 2019 IEEE International Conference on Electro Information Technology (EIT) (pp. 427-433). IEEE.
- Aloysius, N., & Geetha, M. (2017, April). A review on deep convolutional neural networks. In 2017 International Conference on Communication and Signal Processing (ICCSP) (pp. 0588-0592). IEEE. DOI: 10.1109/ICCSP.2017.8286426.
- Al-Turjman, F. M. (2016, November). Towards smart ehealth in the ultra large-scale Internet of Things era. In 2016 23rd Iranian Conference on Biomedical Engineering

- and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME) (pp. 102-105). IEEE.
- Amisha, P. M., Pathania, M., & Rathaur, V. K. (2019). Overview of artificial intelligence in medicine. *Journal of family medicine and primary care*, 8(7), 2328.
- Baltimore D, Berg P, Botchan M, Carroll D, Charo RA, Church G, ... Greely HT (2015) A prudent path forward for genomic engineering and germline gene modification. *Science* 348(6230):36–38. <https://doi.org/10.1126/science.aab1028>
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Morneau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709–1712. <https://doi.org/10.1126/science.1138140>
- Barrangou, R., & Doudna, J. A. (2016). Applications of CRISPR technologies in research and beyond. *Nature biotechnology*, 34(9), 933-941.
- Batista AC, Pacheco LG (2018) Detecting pathogens with zinc-finger, TALE and CRISPR-based programmable nucleic acid binding proteins. *J Microbiol Methods* 152:98–104. <https://doi.org/10.1016/j.mimet.2018.07.024>
- Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45:273–297. <https://doi.org/10.1146/annurevgenet-110410-132430>
- Brownlee J (2016) Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end. *Machine Learning Mastery*
- Bruch R, Baaske J, Chatelle C, Meirich M, Madlener S, Weber W, Dincer C, Urban GA (2019) CRISPR/ Cas13a-powered electrochemical microfluidic biosensor for nucleic acid amplification-free miRNA diagnostics. *Adv Mater* 31:1905311. <https://doi.org/10.1002/adma.201905311>
- Butt, C., Gill, J., Chun, D., & Babu, B. A. (2020). Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence*, 1.
- Chang, R. I., Chiu, Y. H., & Lin, J. W. (2020). Two-stage classification of tuberculosis culture diagnosis using convolutional neural network with transfer learning. *The Journal of Supercomputing*, 1-16.

- Chen, C. M., Chou, Y. H., Tagawa, N., & Do, Y. (2013). Computer-aided detection and diagnosis in medical imaging. *Computational and Mathematical Methods in Medicine*. vol. 2013, Article ID 790608, 2 pages, 2013. doi:10.1155/2013/790608
- Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research*, 42(10), 6091-6105.
- Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. arXiv preprint arXiv:2003.11597.
- Chowdhury, M. E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., ... & Reaz, M. B. I. (2020). Can AI help in screening viral and COVID-19 pneumonia?. arXiv preprint arXiv:2003.13145.
- Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., ... & de Albuquerque, V. H. C. (2020). A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. *Applied Sciences*, 10(2), 559.
- Cicerone, M. T., & Camp Jr, C. H. (2019). Potential roles for spectroscopic coherent Raman imaging for histopathology and biomedicine. In *Neurophotonics and Biomedical Spectroscopy* (pp. 547-570). Elsevier.
- Costa Filho, C. F. F., Levy, P. C., Xavier, C. D. M., Fujimoto, L. B. M., & Costa, M. G. F. (2015). Automatic identification of tuberculosis mycobacterium. *Research on biomedical engineering*, 31(1), 33-43. <http://dx.doi.org/10.1590/2446-4740.0524>
- Das, D., Santosh, K. C., & Pal, U. (2020). Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Physical and engineering sciences in medicine*, 1-11.
- Dawud, A. M., Yurtkan, K., & Oztoprak, H. (2019). Application of Deep Learning in Neuroradiology: Brain Haemorrhage Classification Using Transfer Learning. *Computational Intelligence and Neuroscience*, 2019.
- Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5), 198-211.
- Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* 346(6213):1258096. <https://doi.org/10.1126/science.1258096>

- Druszczyńska, M., Kowalewicz-Kulbat, M., Fol, M., Włodarczyk, M., & RuDnICKA, W. (2012). Latent M. tuberculosis infection--pathogenesis, diagnosis, treatment and prevention strategies. *Pol J Microbiol*, 61(1), 3-10.
- El-Melegy, M., Mohamed, D., El-Melegy, T., & Abdelrahman, M. (2019). Identification of Tuberculosis Bacilli in ZN-Stained Sputum Smear Images: A Deep Learning Approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 0-0).
- Echtioui, A., Zouch, W., Ghorbel, M., Mhiri, C., & Hamam, H. (2020). <? covid19?> Detection Methods of COVID-19. *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, 2472630320962002.
- Falade, A. G., & Ayede, A. I. (2011). Epidemiology, aetiology and management of childhood acute community-acquired pneumonia in developing countries--a review. *African journal of medicine and medical sciences*, 40(4), 293-308.
- Fan, C., & Hauser, H. (2018, June). Fast and accurate cnn-based brushing in scatterplots. In *Computer Graphics Forum* (Vol. 37, No. 3, pp. 111-120).
- Geraldi, A., & Giri-Rachman, E. A. (2018). Synthetic biology-based portable in vitro diagnostic platforms. *Alexandria journal of medicine*, 54(4), 423-428.
- Gilani, Z., Kwong, Y. D., Levine, O. S., Deloria-Knoll, M., Scott, J. A. G., O'Brien, K. L., & Feikin, D. R. (2012). A literature review and survey of childhood pneumonia etiology studies: 2000–2010. *Clinical infectious diseases*, 54(suppl_2), S102-S108.
- Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, ... Morneau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468(7320):67. <https://doi.org/10.1038/nature09523>
- Gilvary, C., Madhukar, N., Elkhader, J., & Elemento, O. (2019). The missing pieces of artificial intelligence in medicine. *Trends in pharmacological sciences*, 40(8), 555-564.
- Gómez P, Semmler M, Schützenberger A, Bohr C, Döllinger M. Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. *Medical & biological engineering & computing*. 2019;57(7):1451-63.

- González-Martín, J., García-García, J. M., Anibarro, L., Vidal, R., Esteban, J., Blanquer, R., ... & Ruiz-Manzano, J. (2010). Consensus document on the diagnosis, treatment and prevention of tuberculosis. *Archivos de Bronconeumología ((English Edition))*, 46(5), 255-274.
- Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J, ... Myhrvold C (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356(6336):438–442. <https://doi.org/10.1126/science.aam9321>
- Gootenberg JS, Abudayyeh OO, Kellner MJ JJ, Collins JJ, Zhang F (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* 360(6387):439–444. <https://doi.org/10.1126/science.aag0179>
- Hajian, R., Balderston, S., Tran, T., DeBoer, T., Etienne, J., Sandhu, M., ... & Paredes, J. (2019). Detection of unamplified target genes via CRISPR–Cas9 immobilized on a graphene field-effect transistor. *Nature biomedical engineering*, 3(6), 427-437.
- Halalli, B., & Makandar, A. (2018). Computer aided diagnosis-medical image analysis techniques. *Breast Imaging*, 85.
- Hall, L. O., Paul, R., Goldgof, D. B., & Goldgof, G. M. (2020). Finding covid-19 from chest x-rays using deep learning on a small dataset. *arXiv preprint arXiv:2004.02060*.
- Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, S36-S40.
- Han, X., Zhong, Y., Cao, L., & Zhang, L. (2017). Pre-trained AlexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sensing*, 9(8), 848. <https://doi.org/10.3390/rs9080848>
- Huang M, Zhou X, Wang H, Xing D (2018) Clustered regularly interspaced short palindromic repeats/ Cas9 triggered isothermal amplification for site-specific nucleic acid detection. *Anal Chem* 90(3):2193– 2200. <https://doi.org/10.1021/acs.analchem.7b04542>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

- Helwan, A., & Abiyev, R. (2016). Shape and texture features for identification of breast cancer. In *Proceedings of the World Congress on Engineering and Computer Science* (Vol. 2, pp. 19-21).
- Helwan, A., & Uzun Ozsahin, D. (2017). Sliding window based machine learning system for the left ventricle localization in MR cardiac images. *Applied Computational Intelligence and Soft Computing*, 2017. <https://doi.org/10.1155/2017/3048181>.
- Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327(5962):167–170. <https://doi.org/10.1126/science.1179555>
- Ishino Y, Krupovic M, Forterre P (2018) History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *J Bacteriol* 200(7):e00580–e00517. <https://doi.org/10.1128/JB.00580-17>
- Kallianos, K., Mongan, J., Antani, S., Henry, T., Taylor, A., Abuya, J., & Kohli, M. (2019). How far have we come? Artificial intelligence for chest radiograph interpretation. *Clinical radiology*.
- Katti, M. K. (2004). Pathogenesis, diagnosis, treatment, and outcome aspects of cerebral tuberculosis. *Medical Science Monitor*, 10(9), RA215-RA229.
- Kermany, Daniel; Zhang, K. G. M. (2018). Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley Data*.
- Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2016). On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*.
- Khan, M. T., Kaushik, A. C., Malik, S. I., Ali, S., & Wei, D. (2019). Artificial neural networks for prediction of tuberculosis disease. *Frontiers in microbiology*, 10, 395. <https://doi.org/10.3389/fmicb.2019.00395>.
- Klassen, V. I., Safin, A. A., Maltsev, A. V., Andrianov, N. G., Morozov, S. P., & Vladzymirskyy, A. V. (2018). AI-based screening of pulmonary tuberculosis: diagnostic accuracy. *Journal of eHealth Technology and Application*, 16(1), 28-32.
- Koo B, Kim DE, Kweon J, Jin CE, Kim SH, Kim Y, Shin Y (2018) CRISPR/dCas9-mediated biosensor for detection of tick-borne diseases. *Sensors Actuators B Chem* 273:316–321. <https://doi.org/10.1016/j.snb.2018.06.069>

- Koonin EV, Makarova KS, Zhang F (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 37:67–78.
<https://doi.org/10.1016/j.mib.2017.05.008>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2), 574-582.
- Lazcka O, Del Campo FJ, Munoz FX (2007) Pathogen detection: a perspective of traditional methods and biosensors. *Biosens Bioelectron* 22(7):1205–1217.
<https://doi.org/10.1016/j.bios.2006.06.036>
- Li Y, Li S, Wang J, Liu G (2019) CRISPR/Cas systems towards next-generation biosensing. *Trends Biotechnol* 37:730–743.
<https://doi.org/10.1016/j.tibtech.2018.12.005>
- Li Y, Liu L, Liu G (2019) CRISPR/Cas multiplexed biosensing: a challenge or an insurmountable obstacle? *Trends Biotechnol* 37:792–795.
<https://doi.org/10.1016/j.tibtech.2019.04.012>
- Liang, J., & Liu, R. (2015, October). Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network. In *2015 8th International Congress on Image and Signal Processing (CISP)* (pp. 697-701). IEEE. DOI: 10.1109/CISP.2015.7407967.
- Liu X, Wu S, Xu J, Sui C, Wei J (2017) Application of CRISPR/Cas9 in plant biology. *Acta Pharm Sin B* 7(3):292–302. <https://doi.org/10.1016/j.apsb.2017.01.002>
- Liz, H., Sánchez-Montañés, M., Tagarro, A., Domínguez-Rodríguez, S., Dagan, R., & Camacho, D. (2020). Ensembles of Convolutional Neural Networks for pediatric pneumonia diagnosis. *arXiv preprint arXiv:2010.02007*.
- Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, ... Xie Y (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. <https://doi.org/10.1016/j.molp.2015.04.007>

- Mathur, S., Fuchs, A., Bielicki, J., Van Den Anker, J., & Sharland, M. (2018). Antibiotic use for community-acquired pneumonia in neonates and children: WHO evidence review. *Paediatrics and international child health*, 38(sup1), S66-S75.
- Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5-6), 555-559. [https://doi.org/10.1016/S0893-6080\(03\)00115-1](https://doi.org/10.1016/S0893-6080(03)00115-1).
- Mei, X., Lee, H. C., Diao, K. Y., Huang, M., Lin, B., Liu, C., ... & Bernheim, A. (2020). Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nature Medicine*, 1-5.
- Men, K., Geng, H., Cheng, C., Zhong, H., Huang, M., Fan, Y., ... & Xiao, Y. (2019). More accurate and efficient segmentation of organs-at-risk in radiotherapy with convolutional neural networks cascades. *Medical physics*, 46(1), 286-292.
- Mnih, A., & Hinton, G. E. (2009). A scalable hierarchical distributed language model. In *Advances in neural information processing systems* (pp. 1081-1088).
- Mintz, Y., & Brodie, R. (2019). Introduction to artificial intelligence in medicine. *Minimally Invasive Therapy & Allied Technologies*, 28(2), 73-81.
- Mohammadi, R., Salehi, M., Ghaffari, H., Rohani, A. A., & Reiazi, R. (2020). Transfer Learning-Based Automatic Detection of Coronavirus Disease 2019 (COVID-19) from Chest X-ray Images. *Journal of Biomedical Physics and Engineering*, 10(5), 559-568.
- Muyama, L., Nakatumba-Nabende, J., & Mudali, D. (2019, December). Automated Detection of Tuberculosis from Sputum Smear Microscopic Images Using Transfer Learning Techniques. In *International Conference on Intelligent Systems Design and Applications* (pp. 59-68). Springer, Cham.
- Myhrvold C, Freije CA, Gootenberg JS, Abudayyeh OO, Metsky HC, Durbin AF, Kellner MJ et al (2018) Field-deployable viral diagnostics using CRISPR-Cas13. *Science* 360(6387):444–448. <https://doi.org/10.1126/science.aas8836>
- Narin, A., Kaya, C., & Pamuk, Z. (2020). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849*.

- Nayak M, Kotian A, Marathe S, Chakravorty D (2009) Detection of microorganisms using biosensors—a smarter way towards detection techniques. *Biosens. Bioelectron* 25(4):661–667. <https://doi.org/10.1016/j.bios.2009.08.037>
- Oyedotun, O. K., Olaniyi, E. O., & Khashman, A. (2017). A simple and practical review of over-fitting in neural network learning. *International Journal of Applied Pattern Recognition*, 4(4), 307-328. <https://doi.org/10.1504/IJAPR.2017.089384>.
- Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, 103792.
- Paiva JS, Cardoso J, Pereira T (2018) Supervised learning methods for pathological arterial pulse wave differentiation: a SVM and neural networks approach. *Int J Med Inform* 109:30–38.
- Panicker, R. O., Kalmady, K. S., Rajan, J., & Sabu, M. K. (2018). Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods. *Biocybernetics and Biomedical Engineering*, 38(3), 691-699.
- Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW, ... Daringer NM (2016) Rapid, lowcost detection of Zika virus using programmable biomolecular components. *Cell* 165(5):1255–1266. <https://doi.org/10.1016/j.cell.2016.04.059>
- Patchesung, M., Jantarug, K., Pattama, A., Aphicho, K., Suraritdechachai, S., Meesawat, P., ... & Athipanyasilp, N. (2020). Clinical validation of a Cas13-based assay for the detection of SARS-CoV-2 RNA. *Nature Biomedical Engineering*, 1-10.
- Polesky, A., Grove, W., & Bhatia, G. (2005). Peripheral tuberculous lymphadenitis: epidemiology, diagnosis, treatment, and outcome. *Medicine*, 84(6), 350-362.
- Prashanth, D. S., Mehta, R. V. K., & Sharma, N. (2020). Classification of Handwritten Devanagari Number—An analysis of Pattern Recognition Tool using Neural Network and CNN. *Procedia Computer Science*, 167, 2445-2457.
- Priya, E., & Srinivasan, S. (2015). Separation of overlapping bacilli in microscopic digital TB images. *Biocybernetics and Biomedical Engineering*, 35(2), 87-99.
- Quan, H., Xu, X., Zheng, T., Li, Z., Zhao, M., & Cui, X. (2020). DenseCapsNet: Detection of COVID-19 X-ray Images Using a Capsule Network.

- Quinn, J. A., Nakasi, R., Mugagga, P. K., Byanyima, P., Lubega, W., & Andama, A. (2016, December). Deep convolutional neural networks for microscopy-based point of care diagnostics. In Machine Learning for Healthcare Conference (pp. 271-281).
- Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. In Advances in neural information processing systems (pp. 3347-3357).
- Rajaraman, S., Candemir, S., Kim, I., Thoma, G., & Antani, S. (2018). Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Applied Sciences*, 8(10), 1715.
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Lungren, M. P. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.
- Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., ... & Kashem, S. (2020). Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray. *Applied Sciences*, 10(9), 3233.
- Rath, D., Amlinger, L., Rath, A., & Lundgren, M. (2015). The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie*, 117, 119-128.
- Ravishankar, H., Sudhakar, P., Venkataramani, R., Thiruvenkadam, S., Annangi, P., Babu, N., & Vaidya, V. (2016). Understanding the mechanisms of deep transfer learning for medical images. In Deep learning and data labeling for medical applications (pp. 188-196). Springer, Cham.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3), 211-252. DOI 10.1007/s11263-015-0816-y.
- Sahlol, A. T., Abd Elaziz, M., Tariq Jamal, A., Damaševičius, R., & Farouk Hassan, O. (2020). A Novel Method for Detection of Tuberculosis in Chest Radiographs Using Artificial Ecosystem-Based Optimisation of Deep Neural Network Features. *Symmetry*, 12(7), 1146.

- Saraiva, A., Ferreira, N., Sousa, L., Carvalho da Costa, N., Sousa, J., Santos, D., & Soares, S. (2019). Classification of Images of Childhood Pneumonia using Convolutional Neural Networks. In 6th International Conference on Bioimaging (pp. 112-119).
- Serte, S., & Demirel, H. (2019). Gabor wavelet-based deep learning for skin lesion classification. *Computers in biology and medicine*, 113, 103423.
- Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., ... & Shi, Y. (2020). Lung infection quantification of covid-19 in ct images with deep learning. *arXiv preprint arXiv:2003.04655*.
- Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5), 1285-1298.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60.
- Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., ... & Zhang, F. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular cell*, 60(3), 385-397.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Smith, K. P., Kang, A. D., & Kirby, J. E. (2018). Automated interpretation of blood culture gram stains by use of a deep convolutional neural network. *Journal of clinical microbiology*, 56(3), e01521-17. DOI: 10.1128/JCM.01521-17.
- Smith, L. N., & Topin, N. (2016). Deep convolutional neural network design patterns. *arXiv preprint arXiv:1611.00847*.
- Stefano, G. B., & Fernandez, EA. (2017). Biosensors: enhancing the natural ability to sense and their dependence on bioinformatics.
<https://doi.org/10.12659/MSM.905800>
- Stephen, O., Sain, M., Maduh, U. J., & Jeong, D. U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of healthcare engineering*, 2019.

- Stewart, G. R., Robertson, B. D., & Young, D. B. (2003). Tuberculosis: a problem with persistence. *Nature Reviews Microbiology*, 1(2), 97-105.
- Tahamtan, A., & Ardebili, A. (2020). Real-time RT-PCR in COVID-19 detection: issues affecting the results.
- Tsai, K. S., Chang, H. L., Chien, S. T., Chen, K. L., Chen, K. H., Mai, M. H., & Chen, K. T. (2013). Childhood tuberculosis: epidemiology, diagnosis, treatment, and vaccination. *Pediatrics & Neonatology*, 54(5), 295-302.
- Turkoglu, M. (2020). COVIDetectionNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble. *Applied Intelligence*, 1-14.
- Wang, D., Wang, X. W., Peng, X. C., Xiang, Y., Song, S. B., Wang, Y. Y., ... & Ma, Z. W. (2018). CRISPR/Cas9 genome editing technology significantly accelerated herpes simplex virus research. *Cancer gene therapy*, 25(5), 93-105.
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).
- World Health Organization, Household Air Pollution and Health [Fact Sheet], WHO, Geneva, Switzerland, 2018, <http://www.who.int/newa-room/fact-sheets/detail/household-airpollution-and-health>
- World Health Organization Report on Tuberculosis. Available at <https://apps.who.int/iris/bitstream/handle/10665/329368/9789241565714-eng.pdf?ua=1>
- Xiong, Y., Ba, X., Hou, A., Zhang, K., Chen, L., & Li, T. (2018). Automatic detection of mycobacterium tuberculosis using artificial intelligence. *Journal of thoracic disease*, 10(3), 1936.
- Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., ... & Chen, Q. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *International Journal of Oral Science*, 12(1), 1-5. <https://doi.org/10.1038/s41368-020-0074-x>.

- Yahiaoui, A., Er, O., & Yumusak, N. (2017). A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines. *Biomedical Research* (0970-938X), 28(9).
- Yang, W., & Yan, F. (2020). Patients with RT-PCR-confirmed COVID-19 and normal chest CT. *Radiology*, 295(2), E3-E3.
- Yu, Y., Lin, H., Meng, J., Wei, X., Guo, H., & Zhao, Z. (2017). Deep transfer learning for modality classification of medical images. *Information*, 8(3), 91.
- Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine*, 15(11).
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., ... & Koonin, E. V. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163(3), 759-771.
- Zhang, F., Abudayyeh, O. O., & Gootenberg, J. S. (2020). A protocol for detection of COVID-19 using CRISPR diagnostics. *A protocol for detection of COVID-19 using CRISPR diagnostics*, 8.
- Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, 42, 146-157.
- Zhang Y, Qian L, Wei W, Wang Y, Wang B, Lin P, ... Cheng S (2016) Paired design of dCas9 as a systematic platform for the detection of featured nucleic acid sequences in pathogenic strains. *ACS Synth Biol* 6(2):211–216.
<https://doi.org/10.1021/acssynbio.6b00215>
- Zulkifley, M. A., Abdani, S. R., & Zulkifley, N. H. (2020). COVID-19 Screening Using a Lightweight Convolutional Neural Network with Generative Adversarial Network Data Augmentation. *Symmetry*, 12(9), 1530.

APPENDICES

APPENDIX 1

MACHINE VS HUMAN 1

Table 1: Results of participants Vs Actual results

S No	Machine	Beginner 1	Certified 1	Beginner 2	Certified 2	Actual (Real) results
1	+	+	+	+	+	+
2	+	+	+	+	+	+
3	+	+	+	+	+	+
4	-	+	+	+	+	+
5	+	-	-	-	-	-
6	-	-	-	-	-	-
7	-	-	-	-	-	-
8	+	+	+	+	+	+
9	-	-	-	-	-	-
10	+	+	+	+	+	+
11	+	+	+	+	+	+
12	+	+	+	+	+	+
13	+	+	+	+	+	+
14	+	+	+	+	+	+
15	-	-	-	-	-	-
16	-	-	-	-	-	-
17	+	-	-	-	-	-
18	+	+	+	+	+	+
19	+	+	+	+	+	+
20	+	+	+	+	+	+
21	-	-	-	-	-	-
22	-	-	-	-	-	-
23	+	+	+	+	+	+
24	+	+	+	+	+	+
25	+	+	+	+	+	+
26	+	+	+	+	+	+
27	+	+	+	+	+	+
28	-	-	-	-	-	-
29	-	-	-	-	-	-
30	+	+	+	+	+	+
31	-	-	-	-	-	-
32	-	-	-	-	-	-
33	+	+	+	+	+	+
34	+	-	-	-	-	-
35	-	-	-	-	-	-
36	+	+	+	+	+	+
37	+	+	+	+	+	+
38	+	+	+	+	+	+
39	-	-	-	-	-	-

40	-	-	-	-	-	-
41	+	+	+	+	+	+
42	+	+	+	+	+	+
43	+	+	+	+	+	+
44	-	-	-	-	-	-
45	+	+	+	+	+	+
46	+	+	+	+	+	+
47	+	+	+	+	+	+
48	+	+	+	+	+	+
49	-	-	-	-	-	-
50	-	+	+	+	+	+
	45 CC	46 CC	49 CC	45 CC	50 CC	
	5 MI	4 MI	1 MI	5 MI	0 MI	

Key:

CC = Correctly classified or identified

MI = Miss-interpreted (highlighted in different colours)

APPENDIX 2

MACHINE VS HUMAN RESULT

Table 2: Distribution of Tuberculosis positive and negative in confusion matrix for 50 images

Participants	True Positive	False Positive	False Negative	True Negative
Machine	31	3	2	19
Beginner 1	31	0	4	19
Beginner 2	31	0	5	19
Certified 1	31	0	1	19
Certified 2	31	0	0	19

APPENDIX 3 MACHINE VS HUMAN 2

Table 3 Results of participants Vs Actual results for 30 Images

s/No	Machine	Beginner 1	Certified 1	Beginner 2	Certified 2	Actual results
1	-	+	-	-	-	-
2	-	-	-	-	-	-
3	+	+	+	+	+	+
4	+	+	+	+	+	+
5	-	-	-	-	-	-
6	+	+	+	+	+	+
7	-	-	-	-	-	-
8	-	-	-	-	-	-
9	+	+	+	+	+	+
10	+	+	+	+	+	+
11	+	+	-	-	+	+
12	+	+	+	+	+	+
13	-	-	-	-	-	-
14	+	+	+	+	+	+
15	+	+	+	+	+	+
16	-	-	-	-	-	-
17	+	+	+	+	+	+
18	-	-	-	-	-	-
19	-	+	-	-	+	-
20	-	-	-	-	-	-
21	+	+	+	+	+	+
22	+	+	+	+	+	+
23	+	+	+	+	+	+
24	+	+	+	+	+	+
25	+	+	+	+	+	+
26	-	-	-	-	-	-
27	+	+	+	+	+	+
28	-	-	-	-	-	-
29	+	+	+	+	+	+
30	+	+	+	+	+	+
		28 CC 2 MI	29 CC 1 MI	28 CC 2 MI	29 CC 1 MI	

Key:

CC = Correctly classified or identified

MI = Miss-interpreted (highlighted in different colours)

APPENDIX 4

MACHINE VS HUMAN 2 RESULT

Table 4: Distribution of Tuberculosis positive and negative in confusion matrix

Participants	True Positive	False Positive	False Negative	True Negative
Machine	18	0	0	12
Beginner 1	18	2	0	12
Beginner 2	18	1	1	12
Certified 1	18	0	1	12
Certified 2	18	1	0	12

APPENDIX 5

CODE

TRAINING

```
>> imds = imageDatastore('training', ...
    'IncludeSubfolders',true, ...
    'LabelSource','foldernames');
>> [imdsTrain,imdsValidation] = splitEachLabel(imds,0.7,'randomized');
>> numTrainImages = numel(imdsTrain.Labels);
>> net = alexnet;
>> net.Layers
>> inputSize = net.Layers(1).InputSize
>> layersTransfer = net.Layers(1:end-3);
>> numClasses = numel(categories(imdsTrain.Labels))
>> layers = [
    layersTransfer

    fullyConnectedLayer(numClasses,'WeightLearnRateFactor',20,'BiasLearnRateFactor',20)
    softmaxLayer
    classificationLayer];
>> pixelRange = [-30 30];
imageAugmenter = imageDataAugmenter( ...
    'RandXReflection',true, ...
    'RandXTranslation',pixelRange, ...
    'RandYTranslation',pixelRange);
augimdsTrain = augmentedImageDatastore(inputSize(1:2),imdsTrain, ...
    'DataAugmentation',imageAugmenter);
>> augimdsValidation = augmentedImageDatastore(inputSize(1:2),imdsValidation);
>> options = trainingOptions('sgdm', ...
    'MiniBatchSize',10, ...
    'MaxEpochs',20, ...
    'InitialLearnRate',1e-4, ...
```

```

'ValidationData',augimdsValidation, ...
'ValidationFrequency',3, ...
'ValidationPatience',Inf, ...
'Verbose',true, ...
'Plots','training-progress');
>> netTransfer = trainNetwork(augimdsTrain,layers,options);

```

VALIDATION

```

>> [YPred,scores] = classify(netTransfer,augimdsValidation);
>> YValidation = imdsValidation.Labels;
>> accuracy = mean(YPred == YValidation)

```

Specificity and Sensitivity

```

>> y = grp2idx(YValidation);
>> test = grp2idx(YPred);
>> classperf(y,test)

```

TESTING

```

>> imdsTest = imageDatastore('Testing', ...
'IncludeSubfolders',true, ...
'LabelSource','foldernames');
>> [YPred,scores] = classify(netTransfer,imdsTest);
>> YValidation = imdsTest.Labels;
>> accuracy = mean(YPred == YValidation)

```

APPENDIX III: ETHICAL APPROVAL DOCUMENT



ETHICAL APPROVAL DOCUMENT

Date:10.02.2021

To the Graduate School of Applied Sciences

For the thesis project entitled as "Application of Artificial Intelligence in Microbiology and Crispr.", the researchers declare that they did not collect any data from human/animal or any other subjects. Therefore, this project does not need to go through the ethics committee evaluation.

Title: PROF. DR.

Name Surname: MEHMET ÖZSÖZ

Signature:

A handwritten signature in blue ink, appearing to be "M. Özsoz".

Role in the Research Project: Supervisor

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name : Abdullahi Umar Ibrahim
Nationality : Nigerian
Date and Place of Birth : 25 October 1987, Zaria.
Marital Status : Single



EDUCATION

a. Schools Attended with Dates

- ❖ Near East University 2017-
- ❖ Cyprus International University 2014-2016
- ❖ Nigerian Institute of science and leather technology, Zaria 2008-2012
- ❖ Science Secondary School Kufaina Zaria, 2002-2008
- ❖ Primary Nuhu Bamalli polytechnic staff School 1996-2002

b. Qualifications Obtained with Dates

- PhD in view 2017-
- ❖ Masters in Bioengineering Feb, 2016
- ❖ National Youth Service corps
(Certificate of National Service) Jun, 2014
- ❖ Higher National Diploma in science Laboratory technology
(Biochemistry option) (Second Class Honors – Division) Dec, 2012
- ❖ National Diploma in science Laboratory technology
(Second Class Honors – Division) Dec, 2010
- ❖ National Examination Council (NECO), July, 2008
- ❖ First School Leaving Certificate (FSLC) June, 2002

c. Research Project/ Seminar

- ❖ Preservation of Tiger nut using Chemical and Natural Antimicrobial agents
- ❖ Higher national Diploma “Comparative study of Sodium Nitrate and Sodium sulphite in g meet Preservation”

- ❖ National Diploma “Investigation of blood Glucose Level in a Blood Sample of patient(Adults) (A case Study of The Department of Chemical Pathology Ahmadu bello Teaching University Zaria)”

d. Publications

- ❖ **CRISPR Technology: Advantages, Limitations and Future Direction. Journal of Biomedical and Pharmaceutical Sciences 2018.**
- ❖ **Comparative study of crispr-cas9 and CRISPR interference (CRISPRi). Journal of Biomedical and Pharmaceutical Sciences 2019.**
- ❖ **Utilization of Nanoparticles For Enhanced Delivery Of Crispr-Cas9 In Cancer Treatment. Journal of Biomedical and Pharmaceutical Sciences 2019.**
- ❖ **Application of Crispr Technology for the Generation of Biofuels: A Review. Journal of Fundamentals of Renewable Energy and Applications. 2019.**
- ❖ **Futuristic CRISPR-based biosensing in the cloud and internet of things era: an overview. Multimedia Tools and Applications 2020.**
- ❖ **Abdullahi Umar Ibrahim, Mehmet Ozsoz, Sertan Serte, Fadi Al-Turjman and Polycarp Shizawaliyi Yakoi4 (2020). Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. Cognitive Computation <https://doi.org/10.1007/s12559-020-09787-5>. Received: 19 August 2020 / Accepted: 21 October 2020. Accepted.**
- ❖ **Abdullahi Umar Ibrahim, Emrah Guler, Meryem Guvenir, Prof Kaya Suer, DR Sertan Serte, Mehmet Ozsoz (2021). Automated detection of mycobacterium tuberculosis using Transfer learning. Journal of Infection in Developing Countries (JIDC). Accepted**

e. Membership of Professional Bodies/Affiliations

f. National Association of Science laboratory Student(NASTES)

TITLE/RESPONSIBILITIES HOLD

- ❖ Vice President, National Association Of Science and Technology Student(NASTES) CHELTECH Chapter
- ❖ Social Director, Kaduna State Association (Kadsu) CHELTECH Chapter
- ❖ Laboratory Captain, Science Secondary School Kufaina, Zaria.
- ❖ Press Club President, Science Secondary School Kufaina, Zaria
- ❖ JET club President, Science Secondary School Kufaina, Zaria
- ❖ Assistant Head boy, Nuhu Bamalli Polytechnic Zaria

EMPLOYMENT HISTORY

I assist in the following:

- ❖ Research Assistant, Biomedical Engineering Near East University
- ❖ Laboratory Technologist at Medical Biochemistry, Kaduna state university
- ❖ Student assistant at Cyprus International University
- ❖ Worked in the Hospital Heart to Heart (HIV UNIT), General Hospital Minna,

Niger state

- ❖ Worked in the Hospital Chemical Pathology Department. Ahmadu Bello University teaching Hospital, Zaria

Here I acquired good Technical skills and became aware of good Laboratory practice.

AWARD/PRIZES

- ❖ Youngest sportsman Award, Sport CDS, National Youth service corps, Niger state, Minna, 2014.
- ❖ Youngest Student Award, Nigerian Institute of Leather and Science Technology, 2010

INTERESTS

I spend my spare-time planning, executing, and love to see the end of achievements of such plans at the set time. I love football, running, I also love music, reading, travelling, driving, meeting and making friends.

REFEREES

(Dr.) Ayse Gunay Kibarar

HOD

Biomedical Engineering

Near East University Directorate

Mobile: +905338717451

Email: aysegunay.kibarar@neu.edu.tr

Prof. Dr. Mehmet Ozsoz

Deen of Engineering

Near East University

Cyprus

Mobile: +905338313735

Email: Mehmet.ozsoz@neu.edu.tr

Abdulsalam Ibrahim Shema

Research Assistant

Cyprus International University

Samaru Zariar, Kaduna State

Mobile: +905488325677

Email: shemadaddy@gmail.com

Turnitin Report : Abdullah Umar Ibrahim

<input type="checkbox"/>	AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
<input type="checkbox"/>	Umar Ibrahim	ABSTRACT.docx	0%	—	—		1482135617	30-Dec-2020
<input type="checkbox"/>	Umar Ibrahim	CONCLUSION.docx	0%	—	—		1482135652	30-Dec-2020
<input type="checkbox"/>	Umar Ibrahim	CHAPTER_1.docx	6%	—	—		1482135630	30-Dec-2020
<input type="checkbox"/>	Umar Ibrahim	CHAPTER_5.docx	8%	—	—		1482135642	30-Dec-2020
<input type="checkbox"/>	Umar Ibrahim	CHAPTER_3.docx	9%	—	—		1482135635	30-Dec-2020
<input type="checkbox"/>	Abdullah Umar	total thesis	11%	—	—		1482149596	30-Dec-2020
<input type="checkbox"/>	Umar Ibrahim	CHAPTER_2.docx	12%	—	—		1482135631	30-Dec-2020
<input type="checkbox"/>	Umar Ibrahim	CHAPTER_4.docx	13%	—	—		1482135638	30-Dec-2020

Prof.Dr.Mehmet Özsöz