GAZİ ÜNİVERSİTESİ

1926

DETECTION OF COVID-19 AND NON-COVID-19 VIRAL PNEUMONIA USING DEEP LEARNING TECHNIQUES

Fidel Morris OMOLLO

MASTERS THESIS

INFORMATION SYSTEMS

GAZİ UNIVERSITY

INFORMATICS INSTITUTE

AUGUST 2022

ETHICAL STATEMENT

I hereby declare that in this thesis study I prepared in accordance with thesis writing rules of Gazi University Informatics Institute.

- All data, information and documents presented in this thesis have been obtained within the scope of academic rules and ethical conduct,
- All information, documents, assessments and results have been presented in accordance with scientific ethical conduct and moral rules,
- All material used in this thesis that are not original to this work have been fully cited and referenced,
- No change has been made in the data used,
- The work presented in this thesis is original,

  or else, I admit all loss of rights to be incurred against me.

Signature

Fidel Morris OMOLLO

30/09/2022

DETECTION OF COVID-19 AND NON-COVID-19 VIRAL PNEUMONIA USING DEEP LEARNING TECHNIQUES

(M. Sc. Thesis)

Fidel Morris OMOLLO

GAZİ UNIVERSITY

INFORMATICS INSTITUTE

August 2022

ABSTRACT

Coronavirus disease (COVID-19) has caused millions of people around the globe to lose their lives. The disease has also adversely affected the economic growth of many countries, especially those based in the developing world. Pneumonia is also one of the leading causes of deaths among children and the elderly in the world. Reverse transcription polymerase chain reaction (RT-PCR) is currently the frequently used testing technique for Coronavirus disease. However, RT-PCR is laborious, expensive and needs specialized medical staff to conduct the test. The objective of this study is to evaluate the performance of deep learning techniques in identifying COVID 19 and non-COVID-19 viral pneumonia. Transfer learning (TL) was the deep learning technique used to conduct the experiments and chest X-ray images was the dataset used to train the model. Seven pre-trained Convolutional Neural Networks (CNNs) were trained, validated, and tested. The X-ray image dataset consisted of 10192 Normal, 1345 viral pneumonia, 3616 COVID-19, and 6012 Lung opacity images. The results produced by the seven models were promising, indicating that the AI based disease diagnostic tools can help in the fast and accurate detection of diseases.

Science Code	:	92432
Key Words	:	COVID-19, viral pneumonia, transfer learning, chest x-ray, deep learning
Page Number	:	75
Supervisor	:	Prof. İsmail ŞAHİN

COVID-19 ILE COVİD-19 OLMAYAN VIRAL PNÖMONYAYI AYIRMAK İÇIN DERIN ÖĞRENME TEKNIKLERININ KULLANILMASI.

(Yüksek Lisans Tezi)

Fidel Morris OMOLLO

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Ağustos 2022

ÖZET

1. GİRİŞ

Koronavirüs hastalığı (COVID-19), dünya genelinde milyonlarca insanın acı çekmesine ve yıkıma uğramasına neden olmuştur. Aileler sevdiklerini kaybetmiş, Ekonomiler çöktü ve sağlık tesisleri, kritik durumdaki katlanarak artan sayıda COVID-19 hastası tarafından boğulmuştur (Haleem, Javaid ve Vaishya, 2020). Hastalık oldukça bulaşıcıdır ve şiddetli akut solunum sendromu koronavirüs 2 (SARS-CoV-2) olarak bilinen viral bir mikroorganizmadan kaynaklanır (İrani, 2022). Virüs esas olarak kurbanlarının solunum sisteminin normal işleyişine saldırır ve bozar. Hastalık ilk olarak Çin'de Wuhan hubei eyaleti olarak bilinen bir eyalette keşfedilmiştir (Li, 2020). Daha sonra hızla tüm dünyaya yayıldı ve 11 Mart 2020'de Dünya Sağlık Örgütü (WHO) tarafından küresel bir pandemi olarak ilan edilmiştir (Dünya Sağlık Örgütü [WHO], 2020).

Küresel sağlık kurumları ve ulusal hükümetler, COVID-19 virüsünün yayılmasını engellemek için katı önlemler almaya zorlandı. Sınır ötesi hareket kısıtlandı, halka açık toplanma alanları kapatıldı, aileler evlerine kapatıldı, toplu test girişimleri yapıldı ve COVID-19 bulaşmış mağdurlar karantinaya zorlandı (Mehtar ve diğerleri, 2020). Karantinanın COVID-19 yayılmasını sınırlamak için kullanılan etkili ve başarılı stratejilerden biri olduğu kanıtlanmıştır. Bir hasta COVID-19 pozitif olarak tanımlandığında, negatif çıkana kadar birkaç hafta kendilerini toplumdan izole etmeye zorlanır (Rahaman, 2020). Bu, enfekte olmuş kişilerin bilmeden hastalığı diğer sağlıklı insanlara yayma olasılığını önlemeye yardımcı olmaktadır.

Ters Transkripsiyon Polimeraz Zincir Reaksiyonu (RT-PCR), şu anda COVID-19 tespiti için yaygın olarak kullanılan test tekniğidir (Sajjad, Shahzad & Mahmood, 2020). Bununla birlikte, test yöntemi, bunu yürütmek için uygun şekilde eğitilmiş tıbbi personel gerekmektedir, sonuçların üretilmesi için çok zaman alır ve PCR makinelerinin ayrıca kurulumu pahalı olan özel laboratuvar tesislerinde barındırılması gerekmektedir. Mahmmud'a (2020) göre, özel PCR laboratuvarı tesisi kurmanın maliyeti 15000$ ile 90000$ arasında değişebilir. Test yönteminin yüksek maliyeti, gelişmekte olan dünyadaki birçok ülkenin COVID-19 test çabalarında geride kalmasının ana nedenlerinden biridir. Ayrıca, son raporlar ayrıca RT-PCR COVID-19 test sonuçlarının %29 ile %40 arasında değişen yanlış negatif oranlarına sahip olabileceğini düşündürmektedir (Zhang, 2021). Yüksek yanlış negatif yüzdeler, hastaların yanlışlıkla COVID-19 negatif test etme olasılığının yüksek olduğu, ancak aslında virüs tarafından enfekte oldukları anlamına gelir.

COVID-19 dışı pnömoni, gelişmiş ülkelerde yaşayan yaşlılar ve gelişmekte olan ülkelerde yaşayan çocuklar arasında önde gelen ölüm nedenlerinden biridir (Ruuskanen, Lahti, Jennings & Murdoch, 2011). Hastalıklara viral veya bakteriyel mikroorganizmalar neden olabilir. Bununla birlikte, bu çalışmada odak noktası COVID-19 ve COVID-19 dışı viral pnömoni olacaktır. Her yıl dünya çapında yaklaşık 450 milyon pnömoni vakası bildirilmekte ve tahminen 4 milyon enfekte kurban hastalığa yenik düşmektedir (Ruuskanen ve ark., 2011). En kötü etkilenen nüfus grubu, 75 yaşın üzerindeki yaşlılar ve 5 yaşın altındaki küçük çocuklardır.

Büyük verilerdeki üstel artış, özellikle internetin yaygın kullanımı nedeniyle, son teknoloji, önceden eğitilmiş derin öğrenme modellerinin geliştirilmesiyle sonuçlanmıştır. Son teknoloji modeller genellikle binlerce veya milyonlarca veri parçasına kadar çalışabilen çok büyük bir veri kümesi üzerinde eğitilir. Modeller tarafından büyük eğitim veri setinden öğrenilen modeller daha sonra benzer veya ilgili bir görevi çözmek için kullanılabilmektedir (Yang, Zhang, Tao & Ma, 2020). Bu çalışmanın amacı, son teknoloji CNN modellerinden bazılarının COVID-19 ve COVID-19 dışı viral pnömoninin otomatik tespitindeki performansını değerlendirmektir.

2. İLGİLİ ARAŞTIRMA ÇALIŞMASI

Krittanawong ve ark. (2019), kardiyovasküler tıpta derin öğrenme tekniklerinin uygulanmasını tartışmıştır. Görüntü segmentasyonu artık kardiyak görüntülemenin daha iyi analizi için kullanılabilmektedir. Akıllı saatler gibi giyilebilir cihazlar artık kardiyovasküler ile ilgili verileri toplamak için kullanılabilmektedir. Yakalanan veriler daha sonra derin öğrenme modelleri ve bir hastanın kardiyovasküler sisteminin başarıyla belirlenen sağlık durumu ile değerlendirilebilmektedir. Yazarlar ayrıca, derin öğrenmenin en büyük faydalarından biri olarak tıbbi görüntüleme analizi ve yorumlarının otomasyonunu önermişler. Model eğitimi sırasında büyük miktarda veriye ihtiyaç duyulması, klinik deneylerde DL tekniklerinin uygulanmasındaki zorluk, model tahmininde kara kutu karar verme ve derin öğrenme modellerini tasarlarken standart bir yolun bulunmaması, derin öğrenmenin bazı sınırlamaları olarak belirtilmiştir. tıp alanında öğrenme.

Yapay Zeka (AI) alanındaki günümüzdeki gelişmeler, tıbbi görüntülemenin yorumlanmasını otomatikleştirme ihtiyacını kolaylaştırmıştır. Göğüs röntgenlerinin analizi otomasyon için yaygın olarak önerilmiştir, çünkü göğüs röntgeni cihazları nispeten ucuzdur, göğüs röntgenleri dünya genelinde yaygın olarak kullanılmaktadır ve görüntüleri yorumlayabilen bir radyolog yetiştirmek birkaç yıl almaktadır. Kallianos ve ark., 2019). Dahmani ve ark. (2020), motor engelli bireylerin elektronik göz kontrollü tekerlekli sandalye kullanarak zahmetsizce hareket etmelerine yardımcı olacak sağlam bir sistem önermiştir. Yazarlar, sistemi dört farklı teknik kullanarak geliştirmişler ve test etmişler. Evrişim Sinir Ağlarını (CNN'ler) kullanan teknik en etkili olarak ortaya çıktı. En iyi performans gösteren CNN, %99,3 doğruluk kaydetmiştir. Bu nedenle tekerlekli sandalyedeki hareketleri kontrol eden bakış tahmincisinde kullanılmak üzere en uygun model seçilmiştir.

Kanserin erken teşhisi ve tedavisi, kanser hastaları için hayatta kalma şansını artırmanın en güvenli yollarından biridir. 206 yılında dünya genelinde 2,8 milyondan fazla kadına kanser teşhisi konmuştur. Mamografi, son 20 yılda meme kanseri hastalarını teşhis etmek için yaygın olarak kullanılan bir tekniktir. Bununla birlikte, teknik, yüksek yanlış pozitif oranlar bildirmiştir. Mamografinin olumsuz yan etkileri, meme kanserinin sorunsuz ve doğru bir şekilde tespit edilmesini kolaylaştıran alternatif Bilgisayar Destekli Teşhis (CAD) araçlarının geliştirilmesini gerekli kılmaktadır (Mambou ve ark., 2018). Ah ark.. (2020), Parkinson hastalığı (PD) olan hastaları teşhis etmek için kullanılan on üç katmandan oluşan bir CNN mimarisi geliştirmiştir. PD, beyin motor fonksiyon yeteneğindeki sürekli düşüşten kaynaklanır. Yazarlar, model veri seti olarak hem sağlıklı hem de PD hastalarından alınan elektroensefalogram (EEG) sinyallerini kullanmışlar. Sinir ağları iyi performans göstererek sırasıyla %88.25, %84.71 ve %91.77 doğruluk, duyarlılık ve özgüllük üretmiştir.

Jo, Nho & Saykin (2019), Alzheimer hastalığı (AH) olan hastaların teşhisine yardımcı olmak için derin öğrenme ve bilgisayar vizyonu kullanmanın yollarını öneren çok sayıda yayını gözden geçirmiştir. Nörogörüntüleme teknolojilerindeki gelişmeler, çok kanallı nörogörüntüleme veri setinin kullanımının artmasıyla sonuçlanmıştır. İncelenen araştırma makalelerinden bazıları hem geleneksel makine öğrenimi hem de gelişmiş derin öğrenme yöntemlerinin bir karışımını kullanırken, diğerleri saf bir derin öğrenme yaklaşımı kullanmıştır. Hem makine öğrenimi hem de derin öğrenme tekniklerinin bir kombinasyonu, AD'nin tanımlanması ve hafif bilişsel bozulma dönüşümü için sırasıyla %98.8 ve 83,7 doğruluk üretmiştir. Saf derin öğrenme yöntemleri, AD'nin tanımlanması ve hafif dönüşüm için sırasıyla %96 ve %84,2 doğruluklar üretmiştir.

Choi et al. (2021), inme tahminine yardımcı olmak için hem uzun kısa süreli bellek (LSTM) hem de konvolüsyonel sinir ağı (CNN) kullanarak bir derin öğrenme modeli geliştirmiştir. Yaşlılar arasında inme baskındır. İnme kurbanları, hastalığın tekrarlama olasılığı yüksek olduğundan ve ayrıca yüksek ölüm oranına sahip olduğundan sürekli gözlem gerekmektedir. İnme muayenesi için yaygın olarak kullanılan yöntemler arasında BT taramaları ve Manyetik rezonans görüntüleme (MRI) taramaları bulunmaktadır. Bununla birlikte, iki teknik, hacimli ve pahalı makinelerin çalıştırılmasını gerektirir ve aynı zamanda zaman alıcıdır. Mevcut yaklaşımlarla ilişkili bazı sınırlamaların ele alınmasına yardımcı olmak için önerilen teknik, model eğitimi ve tahmini için ham elektroensefalografiyi (EEG) kullanmaktadır. Geliştirilen model, %94'lük mükemmel bir vuruş tahmin doğruluğu kaydetmiştir.

Sıtma, özellikle tropik bölgelerde önde gelen ölüm nedenlerinden biridir. Hastalığı teşhis etmenin geleneksel yolu, hastaların kan yaymalarını test etmektir. Mikroskop yardımı ile yaymada sıtma parazitinin bulaştığı kırmızı kan hücreleri tespit edilebilirse, hasta sıtma pozitif olarak belirlenmektedir. Bu geleneksel test yöntemi verimsizdir ve büyük ölçüde testi yapan kişinin becerilerine, uzmanlığına ve deneyimine dayanmaktadır. İnce kan yaymalarında parazit bulaşmış kırmızı kan hücrelerini otomatik olarak tanımlamak için derin transfer öğrenme modellerini kullanan yeni bir teknik önerilmiştir (Shekar, Revathy & Goud, 2020). Yüzden fazla solunum yolu hastalığı, ortak bir semptom olarak öksürüğü paylaşmaktadır. Öksürükler, solunum probleminin nedenini belirlemek için kullanılabilecek zengin ses sinyallerine sahiptir. Kumar ve ark. (2020), öksürük sinyallerini kullanarak on farklı akciğer hastalığını başarıyla tespit edebilen bir derin öğrenme modeli ortaya çıkarmıştır.

Gao & Qian (2017), hastaların BT görüntülerinde beş farklı Tüberküloz (TB) suşunu tespit etmek ve sınıflandırmak için kullanılan bir CNN mimarisi geliştirmiştir. BT görüntülerine, farklı yükseklik ve genişlik boyutlarında dikdörtgen şekilli parçalara dönüştürülerek görüntü segmentasyonu uygulanmıştır. Görüntüler daha sonra model eğitim aşaması başlamadan önce 30 x 30 piksel boyutunda standartlaştırılmıştır. Model performansını değerlendirmek için 300 veri seti kullanılmıştır. Model, 0.22 Kappa puanı ve %41 doğruluk ile yarışmadaki en yüksek doğruluk değeri oluşturmuştur. Marques, Agarwal & de la Torre Díez (2020), COVID-19 tespitini otomatikleştirmek için özelleştirilmiş bir EfficientNet model mimarisiyle çalışmanın yeni bir yolunu önermiştir. CNN tarafından verilen sonuçları doğrulamak için 10 kat katmanlı çapraz doğrulama kullanılmıştır. İki ana deney türü yürütülmüştür. İlki, sağlıklı ve COVID-19 sınıflarından oluşan ikili sınıflandırmaydı. İkincisi, sağlıklı, COVID-19 ve pnömoni sınıflarını içeren çok sınıflı sınıflandırmaydı. EfficientNet CNN'in eğitimi, test edilmesi ve doğrulanması için veri seti olarak göğüs röntgeni görüntüleri kullanılmıştır. İkili ve çok sınıflı sınıflandırma sırasıyla %99.6 ve %96,7 doğruluk üretmiştir.

Rajaraman ve Antani'ye (2020) göre ters transkripsiyon-polimeraz zincir reaksiyonu (RT-PCR), COVID-19 tanısında sıklıkla kullanılan bir test yöntemidir. Ancak yöntemin emek yoğun olması, işletim altyapısını kurmanın pahalı olması ve yanlış pozitif sonuçlar vermesi gibi bazı sınırlamaları vardır. Öte yandan, bilgisayarlı tomografi (BT) görüntülerinin kullanımı da çapraz kontaminasyon olaylarına yatkındır. Yazarlar, göğüs röntgenlerinde COVID-19 ile ilgili opasiteleri tanımak için zayıf bir şekilde etiketlenmiş veri seti ve veri büyütmeyi kullanma yaklaşımı önermişler.

Depresyon aynı zamanda dünya genelinde farklı yaşlardaki insanları olumsuz yönde etkileyen önde gelen zihinsel bozukluklardan biridir. Saidi, Othman & Saoud (2020), depresyondan mustarip bir hastayı otomatik olarak tespit edebilen hibrit bir CNN ve Destek Vektör Makinesi (SVM) makinesi ile ortaya çıkmış. Hibrit model, saf bir CNN tarafından verilen doğruluk sonuçlarından yaklaşık %10 daha fazla olan %68'lik bir doğruluk puanı üretmiştir. Rahman ve ark. (2021), akciğer segmentasyonu ve görüntü iyileştirmenin COVID-19 tespiti için kullanılan göğüs röntgeni veri seti üzerindeki etkilerini incelemiş.

Yazarlar, gama düzeltme, görüntü tamamlayıcı ve histogram eşitleme (HE) gibi farklı görüntü iyileştirme yöntemlerini analiz etmişler. Araştırma deneyinde altı önceden eğitilmiş CNN ve tek bir sığ CNN modeli kullanılmıştır. Gama düzeltme tekniği en iyi performans ölçümlerini verdi. Bununla birlikte, bölümlere ayrılmış veri kümesi kullanıldığında, modeller sırasıyla %95,1, %94,6, %94,6, %94,5 ve %95,6 doğruluk, kesinlik, duyarlılık, F1 puanı ve özgüllük ile önemli ölçüde üstün çıktı üretmiştir.

3. METODOLOJİ

Bu bölüm, araştırma deneylerini yürütmek için kullanılan metodolojiyi ayrıntılı olarak tartışmaktadır. Bu bölüm, derin öğrenme tekniklerini, araştırmada kullanılan derin öğrenme modellerini seçmek için kullanılan kriterleri, model performans ölçümlerini ve çalışma veri kümesini vurgular.

3.1. Derin öğrenme (DL)

Derin öğrenme (DL), bir veri kümesinden gizli kalıpları ve özellikleri öğrenmek için hiyerarşik mimarilere sahip sinir ağlarını kullanan makine öğreniminin bir alt da.lı anlamına gelir (Wooldridge, 2020). Sinir ağları, insan beyninin işleyişini taklit etmeye çalışır. Bilgisayar işlem gücündeki ilerleme ve düşük maliyetli bilgisayar donanımı kaynaklarının üretimi, derin öğrenmenin yükselişine katkıda bulunmaktadır. Ayrıca, derin sinir ağları, temel olarak doğrusal veri kümelerine uygulanan geleneksel makine öğrenimi modellerinin aksine, hem doğrusal olmayan hem de doğrusal veri kümelerinden özellikleri etkili bir şekilde öğrenebilmektedir (Mallick, Dhara & Rath, 2021). DL ile ilgili en büyük zorluklar, büyük bir eğitim veri kümesine duyulan ihtiyaçtır ve ayrıca çoğu durumda DL modellerinin eğitimi zaman alıcıdır ve önemli ölçüde güçlü işleme kapasitesine sahip bilgisayarlar gerekmektedir (Wooldridge, 2020).

Derin öğrenme, makine öğreniminin (ML) bir alt dalı olmasına rağmen, kendisini geleneksel ML'den iki ana şekilde ayırır. Birincisi, derin sinir ağlarını eğitmek ve test etmek için kullanılan veri setinin doğasıdır. Geleneksel makine öğrenimi algoritmaları, öğrenmek ve tahminlerde bulunmak için öncelikle yapılandırılmış, iyi etiketlenmiş veri kümelerini kullanmaktadır (Kaur, 2022). Veri kümesinin yapılandırılmamış olduğu durumlarda, veri kümesini yapılandırılmış bir biçime dönüştürmek için önce ön işleme uygulanmalıdır. Derin öğrenme, ML ile ilişkili birçok veri ön işleme adımını önemli ölçüde en aza indirir. Derin sinir ağları, görüntüler ve metin gibi yapılandırılmamış verilerden özellikleri ve diğer gizli kalıpları otomatik olarak çıkarabilmektedir. ML ve DL arasındaki ikinci fark, eğitim veri kümesinin boyutundadır. DL modelleri, verilerden anlamlı bilgileri başarıyla eğitmek ve öğrenmek için genellikle daha büyük bir veri kümesi gerekmektedir. Ayrıca, DL modellerinin eğitilmesi daha kısa süren ve daha küçük boyutta eğitim verisi gerektirebilen çoğu geleneksel ML modelinden farklı olarak daha uzun sürmektedir (Kaur, 2022).

3.2. Evrişim sinir ağı (CNN)

CNN, bilgisayarla görme görevleri için en popüler kullanılan sinir ağıdır. Evrişim katmanları, tam bağlantılı veya yoğun katmanlar ve havuzlama katmanları olmak üzere üç temel katman bir CNN mimarisini oluşturur (Zhang, Tian & Li, 2018). Üç katmanın her birinin kendi özel işlevi vardır. Bir CNN iki aşamada eğitilir. İleri aşama olarak bilinen ilki, mevcut katman parametrelerini kullanarak giriş verilerini temsil etmeyi gerekmektedir. Daha sonra çıktı (sınıflandırma) katmanı tarafından bir tahmin yapılır ve kesin doğruluk etiketleri ile karşılaştırma yapılarak tahmin hata boyutu belirlenmiştir. Geri faz olarak bilinen ikinci aşama, her bir katman parametresinin gradyanını hesaplamayı ve ağırlıkları ve önyargıları buna göre ayarlamayı içermektedir (Lee ve ark., 2019).

3.3. Transfer öğrenimi

Yeni bir görevle uğraşırken yeterli veri toplamak zor bir görev olabilir. Ayrıca, küçük boyutlu bir veri seti kullanarak mükemmel model sonuçları üretmek çok zor olabilir veya bazı durumlarda ulaşılamaz olabilir. Ancak transfer öğrenimi, modellerimizi eğitmek için küçük bir veri seti kullanmamıza izin vererek ancak yine de mükemmel model ölçümleri elde etmemize izin vererek bu sınırlamayı ele almamıza yardımcı olur (Lu, Behbood, Hao, Zuo, Xue & Zhang, 2015). Transfer öğrenimi, önceden eğitilmiş bir modeli, yeni ancak ilgili bir görev üzerinde başka bir modeli eğitmek için temel olarak yeniden kullanma süreci olarak tanımlanabilmektedir (Day ve Khoshgoftaar, 2017). Transfer öğrenmede, önceden eğitilmiş model ile yeni oluşturulan model arasında genelleştirilmiş bilginin paylaşımı mümkündür, çünkü modeller benzer görevleri yerine getirmeyi amaçlar. Aktarım öğrenimi, zaman ve veri kümesi boyutu gibi modellerimizi oluşturmak ve eğitmek için gereken kaynakları en aza indirmemize yardımcı olur.

Son yıllarda, çeşitli alanlarda ve sektörlerde karmaşık sorunların çözülmesine yardımcı olmak için son teknoloji ürünü birkaç derin öğrenme sistemi geliştirilmiştir. Çoğu durumda, son teknoloji ürünü derin sinir ağlarını oluşturmaktan sorumlu araştırma ekipleri ve bireyler, yapay zeka alanındaki araştırmaları desteklemek için sinir ağları hakkındaki bilgileri halkla paylaşmaktadır. Bu önceden eğitilmiş sinir ağları, derin öğrenme açısından transfer öğrenmenin bel kemiğidir. Derin öğrenmede transfer öğrenme uygulaması, derin transfer öğrenme olarak da adlandırılabilmektedir (Zhang, Zhao & Wu, 2020). Bu yazıda, CNN modellerini eğitmek için yaygın olarak kullanılan derin transfer öğrenme stratejilerinden ikisi olan Özellik çıkarma ve İnce ayar transfer öğrenmesi kullanılmıştır.

Transfer öğrenimi altı ana aşamadan oluşur. İlk aşama, önceden eğitilmiş hazır bir modelin seçimidir. Farklı görevler için uygun geliştirilmiş birkaç model vardır. Bu nedenle, model seçimimize, modelin gerçekleştirmesini istediğimiz görevlerin doğası rehberlik etmelidir. Transfer öğrenmenin etkili bir şekilde yürütülmesi için hedef alan ile önceden eğitilmiş modelin sahip olduğu bilgi arasında güçlü bir karşılıklı bağımlılık olmalıdır (Peirelinck ve ark., 2022). İkinci aşama, bir temel modelin oluşturulmasını içerir. Bu aşamada, çıplak bir sinir ağı mimarisi kullanarak modelimizi sıfırdan eğitmeye veya model eğitim süremizi büyük ölçüde kısaltmamıza yardımcı olacak önceden eğitilmiş model ağ ağırlıklarını indirmeye karar verebilmektedir. Üçüncü aşamada, önceden eğitilmiş sinir ağının eğitimi başlamadan önce, modelin altındaki katmanlar ilk olarak dondurulacaktır. Katmanların dondurulması, önceden eğitilmiş modelin tüm öğrenmesini kaybetmemesini sağlayacaktır (Guérin, Thiery, Nyiri, Gibaru & Boots, 2021). Öğrenmesini kaybeden bir modelin sıfırdan eğitilmesi gerekecek ve bu daha fazla zaman ve hesaplama kaynağı gerektirecektir.

Dördüncü aşamada, donmuş taban katmanlarının üzerine yeni eğitilebilir katman(lar) eklenmektedir. Yürütmek istediğimiz görevin doğasına bağlı olarak tek bir çıktı katmanı veya birkaç eğitilebilir katman eklemeye karar verebilmektedir (Peirelinck ve diğerleri, 2022). Eğitilebilir katmanlar, sinir ağının girdi verilerimizle ilişkili daha ince taneli kalıpları öğrenmesini sağlamaktadır. Nihai çıktı katmanının, sınıflandırma görevleri için kullanılan tam bağlantılı yoğun bir katman olduğu bir durumda. Çıktı yoğun katmanın içerdiği düğüm sayısı, veri kümemizdeki sınıf sayısıyla çakışmalıdır (Guo ve ark., 2016). Örneğin, sinir ağı verileri dört sınıfa sınıflandırmak için eğitilmişse, çıktı tam bağlantılı yoğun katmanın da dört düğümü olmalıdır. Beşinci aşama, gerçek model eğitimini içerir. Oluşturulan model sonuçları, bir dizi deney ve hiper parametre ayarlaması yoluyla daha da geliştirilebilir (Mutasa, Sun & Ha, 2020). Hiper parametre örnekleri arasında öğrenme oranı, parti boyutu, dönem sayısı, momentum, düzenlileştirme sabiti vb. bulunmaktadır. Altıncı adım genellikle isteğe bağlıdır ve ayrıca modelimizin yürüteceği işlev tarafından belirlenmektedir. Adım, performansını daha da artırmak için sinir ağımıza ince ayar yapmayı gerektirmektedir. Modelleri fazla uydurmaya karşı korumak için ince ayar sırasında küçük öğrenme oranı değerleri kullanılmaktadır.

3.4. Çalışmada kullanılan veri kümesi

Çalışmada CNN modellerini eğitmek için kullanılan göğüs röntgeni görüntüleri, bir grup araştırmacı tarafından dünyanın farklı yerlerinden tıp doktorları ile birlikte oluşturulan açık erişimli bir veri tabanından elde edilmiştir (Chowdhury ve ark., 2020). Veritabanı, halka açık github depoları, farklı araştırma yayınları ve İtalyan Tıbbi ve Girişimsel Radyoloji Derneği (SIRM) gibi çeşitli kaynaklardan göğüs röntgeni görüntüleri toplanarak kurulmuştur. Veri seti, Kaggle web sitesinde COVID-19 Radyografi Veritabanı adı altında barındırılmaktadır (Kaggle, 2022). Toplama sırasında olduğu gibi, veri seti 10192 normal, 1345 viral pnömoni, 3616 COVID-19 ve 6012 akciğer opaklığı görüntüsünden oluşmaktadır.

3.5. Model seçimi

Bu çalışmada, önceden eğitilmiş yedi farklı CNN modeli eğitilmiş, doğrulanmış ve test edilmiştir. Modeller MobileNetSmall, MobileNetLarge, EfficieNetV2B2, ResNet50V2, DenseNet 121, NasNetMobile ve Xception'ı içermektedir. Derinliğin benzer yapıya sahip bir modelin performansını nasıl etkilediğini araştırmak için MobileNetV3'ün iki farklı versiyonu kullanılmıştır. Optimizasyon algoritması olarak Adam AMSGrad varyantı kullanılmıştır. AMSgrad, girdi değişkenlerinin öğrenme oranlarında büyük anlık değişikliklerin meydana gelmesini sınırlayarak Adam algoritmasının yakınsama yeteneklerini artırmaya çalışan stokastik bir optimizasyon tekniğidir (Tran & Phong, 2019).

Modeller, hem tam bağlantılı yoğun katman sınıflandırıcısı hem de SVM sınıflandırıcısı ile Özellik çıkarma transfer öğrenimi kullanılarak eğitilirken, 0,001'lik sabit bir öğrenme oranı kullanılmıştır. CNN modellerinin en iyi şekilde yakınsamasını sağlamak ve ayrıca modellerin fazla takılmasını önlemek için İnce ayar transfer öğrenimi kullanılarak model eğitilirken biraz daha düşük bir öğrenme oranı olan 0.0001 kullanılmıştır. Ayrıca, hem sınıflandırıcı olarak yoğun bir katmana sahip öznitelik çıkarma transfer öğrenmesinde hem de model aşırı uydurma caydırıcı olarak ince ayar transfer öğrenmesinde 0,4 oranında bir bırakma katmanı da kullanılmıştır. CNN'leri bir SVM sınıflandırıcı ile özellik çıkarımı kullanarak eğitirken, bırakma katmanı kullanılmadı ve bunun yerine modelleri aşırı uydurmaya karşı korumak için 0.0001 düzenlileştirme faktörüne sahip bir L2 düzenleyici kullanılmıştır. Eğitim sürecinde kullanılan tüm hiper parametre değerleri birkaç deneyden sonra belirlenmiştir.

Mimariye küresel bir ortalama havuzlama katmanı da eklenmiştir. Katmanın işlevi, her bir özellik haritasının ortalamasını hesaplamak ve sınıflandırma için son çıktı katmanına iletilen bir vektör çıktısı almaktır. Bu yazıda, geleneksel tam bağlı yoğun katmanlar yerine küresel ortalama havuzlama katmanları, evrişim yapısıyla doğal olarak uyumlu oldukları ve ayrıca aşırı uydurmayı önlemede becerikli olabilecek herhangi bir parametre optimizasyonu gerektirmedikleri için tercih edilmiştir. Sınıflandırma katmanı olarak tam bağlantılı yoğun bir katman kullanıldığında, ikili sınıflandırmada bir sigmoid aktivasyon işlevi ve ikili_crossentropy kayıp işlevi kullanılmıştır. Çok sınıflı sınıflandırma için, yoğun sınıflandırıcılar kategorik_krosentropi kayıp fonksiyonunu ve softmax aktivasyon fonksiyonunu kullanmıştır. Modelleri özellik çıkarma ve bir SVM sınıflandırıcı tekniği kullanarak eğitirken. Son çıktı katmanı, ikili sınıflandırma için bir L2 çekirdek düzenleyici ve bir menteşe kaybı işlevi ve doğrusal etkinleştirme işlevi kullanılarak bir SVM'ye dönüştürülmüştür. Çok sınıflı sınıflandırma için bir kare_hinge kayıp fonksiyonu ve softmax aktivasyon fonksiyonu uygulanmıştır.

3.6. Veri ön işleme

Göğüs röntgeni görüntüleri, CNN modeli gereksinimlerine bağlı olarak yeniden boyutlandırılmıştır. Modellerin çoğu 224 x 224 piksellik bir resim boyutu gerektiriyordu, kullanılan diğer resim boyutları 260 x 260 piksel ve 229 x 229 pikseldi. Daha fazla eğitim veri seti oluşturmak için döndürme tabanlı görüntü büyütme tekniği kullanılmıştır. Ek olarak oluşturulan görüntüler, modelleri eğitmek için kullanılan 4 sınıftaki görüntü sayısının dengelenmesini sağlamıştır. Dengeli bir eğitim veri seti, model tahmin aşaması sırasında her türlü önyargıyı ortadan kaldırır ve yeni bir veri seti verildiğinde modellerin iyi bir şekilde genelleştirilmesini sağlamaktadır. X-ışını görüntüleri, eğitim için modele beslenmeden önce önce büyütülmüş ve karıştırılmıştır. Normal göğüs röntgeni görüntüleri, artırılmayan tek eğitim veri setiydi, çünkü normal sınıf, eğitim veri setinde en fazla sayıda X-ışını görüntüsünü içeriyordu. Artırılmış görüntüler -40 ila 40 derece arasında bir açıyla döndürülmüştür. Test ve doğrulama veri kümesine görüntü büyütme uygulanmadı. X-ray görüntü veri setini %70 eğitim, %20 test ve %10 doğrulama verisine bölmek için rastgele örnekleme kullanılmıştır. Model eğitimi sırasında 32 görüntüden oluşan sabit bir toplu iş boyutu da kullanılmıştır.

3.8. Performans metrikleri

Eğitilmiş CNN modellerinin performansını değerlendirmek için aşağıdaki gibi çeşitli metrikler kullanılmıştır: (a) Hastalık pozitif olarak doğru bir şekilde kategorize edilen X-ışını görüntüleri olan Gerçek pozitif (TP), (b) X- olan Gerçek Negatif (TN). Doğru olarak hastalık negatif veya sağlıklı olarak tanımlanan ışın görüntüleri, (c) Yanlış bir şekilde hastalıklı bir vaka olarak tanımlanan ancak aslında sağlıklı bir vaka olan X-ışını görüntüleri olan Yanlış Pozitif (FP), (d) X-ışını olan Yanlış Negatif görüntüler yanlış bir şekilde sağlıklı bir vaka olarak sınıflandırılır, ancak aslında olumlu bir vakadır. Ölçümler, CNN modellerinin doğruluğunu, geri çağrılmasını ve kesinliğini ve F1 puanını hesaplamak için kullanılmıştır. Yüksek bir geri çağırma değeri, Yanlış Negatiflerin sayısının düşük olacağı anlamına gelmektedir. Düşük Yanlış Negatif değerler, yanlış şekilde sağlıklı olarak teşhis edilen COVID-19 veya viral pnömoni pozitif hasta sayısının en aza indirilmesini sağlamaktadır. Yüksek hassasiyet, Yanlış Pozitiflerin sayısının düşük olacağı anlamına gelmektedir. Düşük Yanlış Pozitif değerler, COVID-19 veya viral pnömoni ile yanlış teşhis edilen sağlıklı bireylerin sayısının en aza indirilmesini garanti etmektedir.

4. BULGULAR VE TARTIŞMA

Bu çalışmada, iki ana kategorizasyon türü değerlendirilmiştir. İlki, normale karşı viral pnömoni ve normale karşı COVID-19 sınıflarından oluşan ikili sınıflandırmaydı. İkincisi, normale karşı viral pnömoniye ve COVID-19'a karşı ve normale karşı viral pnömoniye karşı COVID-19'a karşı akciğer opaklığından oluşan çok sınıflı sınıflandırmaydı. Akciğer opasite sınıfı, ne COVID-19 ne de viral pnömoni olmayan diğer akciğer enfeksiyonlarını ifade eder. Modeller ayrıca metodoloji bölümünde tartışılan üç farklı derin transfer öğrenme stratejisi kullanılarak eğitilmiştir. İlk strateji, öznitelik çıkarıcı olarak önceden eğitilmiş hazır sinir ağlarını ve son sınıflandırma katmanı olarak tam bağlantılı yoğun bir katmanı kullanmaktı. İkinci strateji, sinir ağları için son çıktı katmanı olarak bir destek vektör makinesinin kullanılması dışında, birincisine benzer. Üçüncü teknik, performanslarını artırmak amacıyla önceden eğitilmiş sinir ağlarının ince ayarını gerektiriyormuştur.

4.1. Özellik çıkarıcı olarak önceden eğitilmiş modeller ve sınıflandırıcı olarak tam bağlantılı katmanlar

Bu model eğitimi ve sınıflandırma kategorisinde, önceden eğitilmiş modellerin üst sınıflandırma katmanı hariç tüm alt katmanlar dondurulur. Bu üst çıktı katmanı daha sonra modelin sınıflandıracağı sınıf sayısına karşılık gelen düğüm sayısını içeren tamamen bağlı başka bir katmanla değiştirildi.

4.1.1. İkili sınıflandırma: normal vs COVID-19

Modellerin ürettiği sonuçlar umut vericiydi. Önceden eğitilmiş CNN'ler, ≥ %91 doğruluk değerleri, ≥ %91 kesinlik değerleri, ≥ %86 geri çağırma değerleri ve ≥ %88 F1 puanları üretmiştir. EfficientNetV2B2 sinir ağı, sırasıyla %99, %98, %98, %98 doğruluk, kesinlik, geri çağırma ve F1-Skoru ile en iyi performansı göstermiştir. MobileNetV3Small ve MobileNetV3Large, sırasıyla 200 saniye ve 204 saniyelik en iyi ve neredeyse aynı eğitim süresini kaydetmiştir. ResNet50V2 ve DesnseNet121, yalnızca bir saniyelik bir farkla sırasıyla 491 saniye ve 492 saniyelik karşılaştırılabilir eğitim süresine sahiptir. Xception sinir ağı, 714 saniyelik en kötü eğitim süresine sahiptir.

4.1.2. İkili sınıflandırma: normal vs viral pnömoni

Tüm CNN'ler doğruluk puanları ≥ %97, kesinlik puanları ≥ %95, hatırlama puanları ≥ %92 ve F1-Skorları ≥ %93 üretmiştir. EffcientNetV2B2, sırasıyla %99, %98, %98, %98 doğruluk, kesinlik, hatırlama ve F1-Skoru ile en iyi performans gösteren CNN olmuştur. NASMobile %97 doğruluk, %95 hassasiyet, %92 hatırlama ve %93 F1-Skoru ile dört ölçümün tamamında en düşük performans değerlerini kaydetmiştir. 190 saniye ile en kısa eğitim süresine sahip MobileNetV3Small'ı ve 197 saniyelik eğitim süresi kaydeden MobileNetV3Large yakından takip etmiştir. Xception, 716 saniye ile en uzun eğitim süresine sahiptir. Olağanüstü model performans metriklerine rağmen EfficientNetV2B2, 548 saniye ile en uzun ikinci eğitim süresine sahiptir.

4.1.3. Çok sınıflı sınıflandırma: normal vs viral pnömoni ve COVID-19

Sinir ağlarının her birinin doğruluk değeri ≥ %90 idi. Her modelin kesinliği, hatırlaması ve F1 puanı ≥ %89'du. En iyi performans gösteren CNN, doğruluk, kesinlik, geri çağırma ve ≥ %95 F1-skor değerleri üreten MobileNetV3Large oldu. İkinci en iyi performans gösteren model, doğruluk, kesinlik, geri çağırma ve F1-Skor değerlerini ≥ %94 kaydeden EfficeintNetV2B2 idi. MobileNetV3Small 275 saniye ile en kısa süreyi kaydederken, onu 283 saniyelik eğitim süresiyle MobileNetV3Large yakından takip etmiştir. Xception, 1055 saniye ile en uzun eğitim süresine sahiptir.

4.1.4. Çok sınıflı sınıflandırma: normal vs viral pnömoni vs COVID-19 vs akciğer opaklığı

Önceden eğitilmiş modellerin her birinin doğruluk, hatırlama, kesinlik ve F1-Skor değeri ≥ %81 idi. Bu sınıflandırma kategorisinde, EffcientNetV2B2 ve MobileNetV3Large, neredeyse benzer performans sonuçları üretmiştir. İki sinir ağı, %89'luk aynı doğruluk puanına sahiptir. Ancak EfficientNetV2B2, %1 oranında biraz daha iyi bir F1 Puanına sahiptir. Xception 1380 saniye ile en kötü eğitim süresine sahip olurken, onu 1033 saniye ile EfficientNetV2B2 takip etmiştir. MobileNetV3Small ve MobileNetV3Large, iki model arasında 10 saniyelik farkla en iyi eğitim süresine sahiptir. MobileNetV3Small 362 saniyelik bir eğitim süresi kaydetti ve MobileNetV3Large 372 saniyelik bir eğitim süresi kaydetmiştir.

4.2. Özellik çıkarıcı olarak önceden eğitilmiş modeller ve sınıflandırıcı olarak destek vektör makinesi (SVM)

Bu model eğitimi ve sınıflandırma kategorisinde, en üstteki sınıflandırma katmanı hariç, önceden eğitilmiş modelin tüm katmanları dondurulur. Önceden eğitilmiş modelin en üst katmanı, sınıflandırma görevini gerçekleştiren bir destek vektör makinesi ile değiştirilirken, donmuş katmanlar özellik çıkarıcı görevi görür.

4.2.1. İkili sınıflandırma: normal vs COVID-19

Önceden eğitilmiş modellerin her biri doğruluk değeri ≥ %93, kesinlik değeri ≥ %94, geri çağırma değeri ≥ %87 ve F1 puanı ≥ %90 kaydetmiştir. EfficientNetV2B2 sinir ağı, sırasıyla %97, %97, %94, %95 doğruluk, kesinlik, geri çağırma ve F1-Skoru ile en iyi performans gösteren modeldi. MobileNetV3Small, 187 saniye ile eğitilmesi en az süreyi almıştır. MobileNetV3Large, 192 saniye ile en kısa ikinci eğitim süresini kaydetmıştır. Xception modeli 701 saniye ile en uzun antrenman süresini almıştır.

4.2.2. İkili sınıflandırma: normal vs viral pnömoni

Tüm CNN'ler doğruluk ve kesinlik puanları ≥ %98, hatırlama puanları ≥ %90 ve F1-Skorları ≥ %94 üretti. EffcientNetV2B2, MobileNetV3Small, MobileNetV3Large ve ResNet50V2 neredeyse benzer metrik puanları kaydetmiştir. Dört modelin tümü, %99'luk aynı doğruluk puanına sahipti, tek fark, maksimum %2 ile farklılık gösteren kesinlik, geri çağırma ve F1-Skor değerlerindeydi. MobileNetV3Small, 186 saniye ile en az eğitim süresine sahiptir. Xception, 700 saniyelik en uzun eğitim süresine sahiptir.

4.2.3. Çok sınıflı sınıflandırma: normal vs viral pnömoni vs COVID-19

CNN'ler tarafından verilen doğruluk, kesinlik, hatırlama ve F1-Skor değerleri ≥ %91 idi. MobileNetV3Large, MobileNetV3Small ve EfficientNetV2B2 en iyi performans gösteren CNN'lerdi. Ancak MobileNetV3Large modeli, diğer iki modelden biraz daha iyi bir doğruluk ve F1-Skor değeri, yani %1 daha yüksek kaydetmiştir. MobileNetV3Small, 277 saniye ile en kısa sürede eğitim almıştır. Xception modeli, 1043 saniye ile en uzun antrenman süresini almıştır.

4.2.4. Çok sınıflı sınıflandırma: normal vs viral pnömoni vs COVID-19 vs akciğer opaklığı

Önceden eğitilmiş modellerin her birinin doğruluk, geri çağırma, kesinlik ve F1-Skor değerleri ≥ %81 idi. EffcientNetV2B2 ve MobileNetV3Large, birbirine yakın performans sonuçları üretmiştir. Geri çağırma ve kesinlik ölçümlerinde yalnızca %1'lik bir fark varmış. EffcientNetV2B2 en yüksek kesinlik değerine sahipken MobileNetV3Large en yüksek geri çağırma değerine sahiptir. Xception ve EfficientNetV2B2, sırasıyla 1035 saniye ve 1394 saniye ile en uzun eğitim süresine sahiptir. MobileNetV3Small, 362 saniye ile en az eğitim süresine sahiptir.

4.3. Modelin ince ayarı

Bu model eğitimi ve sınıflandırma kategorisinde, önceden eğitilmiş modelin daha derin katmanları dondurulurken, sığ model katmanlarından bazıları model eğitimi sırasında çözülür. Dondurulmamış katmanların ağırlıkları, model eğitilirken ayarlanır ve bu, modelin eğitim veri kümesine özgü özellikleri daha iyi öğrenmesine olanak tanır.

4.3.1. İkili sınıflandırma: normal vs COVID-19

Tüm modeller mükemmel derecede iyi performans göstererek doğruluk ≥ %98, hassasiyet ≥ %97, hatırlama ≥ %97 ve F1 puanları ≥ %97 üretmiştir. Altı model arasında en iyi performans gösteren model, doğruluk, hassasiyet ve F1 puanı %99 olan EfficientNetV2B2 oldu. Xception sinir ağı, 840 saniye ile eğitmek için en uzun süreyi almaktadır. MobileNetV3Small ve MobileNetV3Large, sırasıyla 211 saniye ve 219 saniye ile en düşük eğitim süresini kaydeder.

4.3.2. İkili sınıflandırma: normal vs viral pnömoni

%98 doğruluk oranına sahip Xception hariç tüm modeller %99 doğruluk puanına sahiptir. Modellerin her biri için kesinlik ve geri çağırma değerleri ≥ %97 idi. Modellerin her birinin F1-Skor değeri ≥ %98 idi. EffcientNetV2B2, doğruluk, kesinlik ve F1 puanı gibi performans ölçütlerinin üçünde %99'luk bir değerle zirveye yerleşmiştir. MobileNetV3Small, 193 saniye ile en kısa eğitim süresini kaydetti, ardından 208 saniye ile eğitim süresi kaydeden MobileNetV3Large izlemiştir. Xception, 835 saniye ile en uzun antrenman süresini almıştır.

4.3.3. Çok sınıflı sınıflandırma: normal vs viral pnömoni vs COVID-19

CNN'lerin doğruluğu ≥ %97 idi. Her modelin kesinliği ve F1 puanı ≥ %96 idi. Modellerin her birinin hatırlama değeri ≥ %95 idi. En iyi performans gösteren CNN, %99 doğruluk, geri çağırma ve F1 puanı ve %98 hassasiyet üreten MobileNetV3Large olmuştur. MobileNetV3Small, 303 saniyelik en iyi süreyi kaydetmiştir. Xception sinir ağı, 1241 saniyeyle en uzun süreye sahiptir.

4.3.4. Çok sınıflı sınıflandırma: normal vs viral pnömoni vs COVID-19 vs akciğer opaklığı

Önceden eğitilmiş modellerin her birinin doğruluk değeri ≥ %91 idi. Geri çağırma ve F1 puanı ≥ %92 ve kesinlik ≥ %93 idi. Bu sınıflandırma kategorisinde EffcientNetV2B2, MobileNetV3Large ve DeseNet121 hemen hemen aynı sonuçları vermiştir. Modeller benzer bir hassasiyete, hatırlamaya ve %95 F1 puanına sahiptir. Tek fark, DenseNet121'in diğer iki modelden %1 daha yüksek olan %94 puan aldığı doğruluktaydı. Xception 1644 saniye ile en kötü eğitim süresine sahip olurken, onu 1172 saniye ile EfficientNetV2B2 takip etmiştir. MobileNetV3Small ve MobileNetV3Large sırasıyla 418 ve 421 saniye ile en iyi eğitim süresine sahiptir.

5. SONUÇ

Viral pnömoni ve COVID-19 hastalarının tespitinde önceden eğitilmiş sinir ağları tarafından üretilen sonuçlar çok umut vericiydi. Modellerin çoğu, sınıflandırma katmanı olarak yoğun bir katman kullanılarak özellik çıkarma transferi öğrenimi kullanılarak eğitildiklerinde en düşük performans metrik değerlerini kaydetmişler. Sınıflandırma görevini gerçekleştirmek için destek vektör makinesi kullanıldığında metrik değerlerinde marjinal bir artış olmuştur. İnce ayarlı aktarım öğrenimi, en iyi model performans metrik değerlerini oluşmuştur. Ancak, önceden eğitilmiş modeller, İnce ayar transfer öğrenimi kullanılarak eğitildiklerinde en uzun eğitim süresini almıştır. Örneğin, Normal ve COVID-19 göğüs röntgeni görüntülerinin sınıflandırılmasında. MobileNetV3Small eğitim süresi, yoğun bir sınıflandırma katmanı kullanan Özellik çıkarma için 187 saniye, Özellik çıkarma ve bir SVM sınıflandırıcı için 200 saniye ve İnce Ayar Aktarımı öğrenimi için 211 saniyeydi. Daha hızlı eğitim süresi, Özellik çıkarma aktarımı öğrenimini, belirli bir göreve uygun modelleri belirlemek için hızlı temel deneyler yapmak için uygun hale getirmektedir. Belirli bir görev için en uygun modeli seçtikten sonra, performansını daha da artırmak için modelde ince ayar yapılabilmektedir. İnce ayardaki daha uzun eğitim süresi, eğitim sırasında güncellenecek parametre sayısını artıran bazı model katmanlarının çözülmesine bağlanabilmektedir.

Önceden eğitilmiş modeller, ikiden fazla sınıftan oluşan çok sınıflı sınıflandırmaya göre ikili sınıflandırmada daha iyi performans ölçütlerine sahiptir. Modellerin çoğu, normal ve viral pnömoni göğüs röntgeni görüntülerini ayırt etmede iyi performans göstermiştir. Modellerin çoğu, kullanılan üç Transfer öğrenme tekniğinin hepsinde %99'a varan doğruluk puanlarına ve %92'nin üzerinde F1-Skor değerlerine sahiptir. İkinci en iyi sınıflandırma kategorisi COVID-19'a karşı normal, ardından normale karşı viral pnömoniye karşı COVID-19 ve en kötü performans gösteren sınıflandırma kategorisi normale karşı viral pnömoniye karşı COVID-19 ve akciğer opasitesiydi. Veri kümesi sınıflarının sayısındaki artışla birlikte model performansındaki sürekli düşüş, bir modelin öğrenmesi gereken veriye özgü özelliklerin sayısındaki artışla ilişkilendirilebilir. Bu nedenle, veri kümesi sınıflarındaki bazı modeller örtüştüğünde model yanlış tahmine eğilimli olabilmektedir. Ayrıca, CNN'leri çok sınıflı sınıflandırmada eğitmek için geçen süre, esas olarak eğitim verilerinin boyutundaki artıştan dolayı, onları ikili sınıflandırmada eğitmek için geçen süreden daha fazladır. Ayrıca, modellerin üç transfer öğrenme tekniğinin tümünü kullanarak eğitilmesi sırasında çok az ya da hiç aşırı uyum yaşanmamış. Bu, veri artırma, bırakma ve L2 düzenleme gibi kullanılan sağlam düzenleme tekniklerine bağlanabilmektedir.

Çalışma sonuçları, derin transfer öğrenmenin sıfırdan karmaşık model mimarileri oluşturmaya etkili ve verimli bir alternatif olarak kabul edilebileceğini göstermektedir. Önceden eğitilmiş modeller, minimum zaman ve çaba kullanarak daha iyi sonuçlar üretebilmektedir. Bu araştırma çalışmasından elde edilen istisnai sonuçlar, transfer öğreniminin COVID-19, COVID-19 dışı viral pnömoni ve diğer akciğer enfeksiyonlarının otomatik tespiti için AI araçlarının geliştirilmesinde kullanılabileceğini de göstermektedir. Ayrıca viral pnömoni ve COVID-19, her ikisi de viral hastalıklar olan benzer belirti ve semptomlar sergiler ve kolayca yanlış teşhis edilebilir. Bu nedenle, AI hastalık tespit araçlarının kullanılması, yanlış teşhis vakalarının önlenmesine yardımcı olabilir ve hastalara uygun tedavinin uygulanmasını sağlayabilmektedir. AI tespit araçları, hastalığın hızlı yayılmasını önlemeye yardımcı olabilecek COVID-19 ile enfekte kişilerin erken taranmasına ve tanımlanmasına da yardımcı olabilmektedir.

Eğitim veri setinin sınırlı mevcudiyeti, çalışmadaki sınırlamalardan biriydi, bu nedenle daha fazla eğitim veri seti oluşturmak ve veri setinin dengeli olmasını sağlamak için görüntü büyütmenin kullanılması. Gelecekteki araştırmalarda, modeller eğitilirken orijinal eğitim veri kümesi sayısı artırılabilir. Ayrıca, ince ayar, model parametrelerinin ayarlanması için sınırsız yollar sunar. Bu nedenle, modellerin performansını daha da artırmak için daha fazla hiper parametre ayarı yapılabilmektedir.

Bilim Kodu	:	92432
Anahtar Kelimeler	:	COVID-19, viral pnömoni, transfer öğrenme, göğüs röntgeni, derin öğrenme
Sayfa Adedi	:	75
Danışman	:	Prof. Dr. İsmail ŞAHİN

THANKS

To all who provided me with all kinds of help and sacrifice during my studies; my supervisor, Prof. İsmail ŞAHİN, my parents and friends, I pass my gratitude and appreciation.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

Page

Page

SYMBOLS AND ABBREVIATIONS

The symbols and abbreviations used in this study are presented below along with their explanations.

Abbreviations	Explanations
COVID-19	Coronavirus disease of 2019
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
WHO	World Health Organization
RT-PCR	Reverse Transcription Polymerase Chain Reaction
CT-Scan	Computerised tomography (CT) scan
DL	Deep learning
CNN	Convolution Neural Network
AI	Artificial Intelligence
CAD	Computer Aided Diagnostic
PD	Parkinson disease
EEG	Electroencephalogram
AD	Alzheimer's disease
LSTM	Long short-term memory
MRI	Magnetic resonance imaging
TB	Tuberculosis
SVM	Support Vector Machine
ML	Machine Learning
ANN	Artificial Neural Networks
TP	True positive
TN	True negative
FP	False positive
FN	False negative

1. INTRODUCTION

Coronavirus disease (COVID-19) has caused suffering and devastation of millions of people across the globe. Families have lost loved ones, Economies have crumbled, and health facilities have been overwhelmed by the exponential number of COVID-19 patients in critical condition (Haleem, Javaid, & Vaishya, 2020). The disease is highly contagious and is caused by a viral microorganism known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Irani, 2022). The virus mainly attacks and disrupts the normal functioning of the respiratory system of its victims. The disease was first discovered in China at a province known as Wuhan hubei province (Li, 2020). It later rapidly spread across the world and was declared as a global pandemic by the World Health Organization (WHO) on the 11th of March 2020 (World Health Organization [WHO], 2020).

Global health institutions and national governments were forced to put in place stringent courses of actions to curb the spread of the COVID-19 virus. Cross border movement was restricted, public gathering areas were shut down, families were locked down in their homes, mass testing initiatives were carried out, and COVID-19 infected victims were forced to quarantine (Mehtar et al., 2020). Quarantining has proven to be one of the effective and successful strategies used to limit COVID-19 spread. Once a patient is identified as COVID-19 positive, they are forced to isolate themselves from the public for some few weeks until they test negative (Rahaman, 2020). This helps to prevent the possibility of infected individuals unknowingly propagating the disease to other healthy people.

Reverse Transcription Polymerase Chain Reaction (RT-PCR) is currently the commonly used testing technique for COVID-19 detection (Sajjad, Shahzad & Mahmood, 2020). However, the testing method requires properly trained medical staff to conduct it, it consumes a lot of time before the results are generated, and PCR machines must also be hosted in specialized lab facilities that are expensive to set up. According to Mahmmud (2020), the cost of building the specialized PCR lab facility can range from $15000 to $90000. The high cost of the testing method is one of the main reasons why many countries in the developing world are lagging behind in the COVID-19 testing efforts. Moreover, recent reports also suggest that the RT-PCR COVID-19 test results may have false negative rates ranging from 29% to 40% (Zhang, 2021).

The high false negative percentages mean there is a high probability of patients mistakenly testing COVID-19 negative but they are in fact infected by the virus.

Non-COVID-19 pneumonia is one of the leading causes of deaths among the elderly living in the developed world, and children living in the developing nations (Ruuskanen, Lahti, Jennings & Murdoch, 2011). The diseases can either be caused by viral or bacterial microorganisms. However, in this study the focus will be on COVID-19 and non-COVID-19 viral pneumonia. Approximately 450 million pneumonia cases are reported globally every year and an estimated 4 million infected victims succumb to the illness (Ruuskanen et al., 2011). The worst affected population group are the elderly with an age beyond 75 years and young children with an age beneath 5 years. Clinical examination of the upper and lower respiratory specimen such as sputum, examination of patients' disease symptoms, and analysis of radiographic images such as CT scans and chest X-rays are some of the common methods used to test viral pneumonia (Tilve et al., 2020).

In recent years, Computer Aided Diagnostic tools have increasingly been used to help in disease identification. Artificial Intelligence is the engine that drives the automated disease diagnostic tools infrastructure (Panesar, 2019: 12 -14). The application of deep learning in the analysis of medical images such as chest X-rays, and CT-scan has helped in early detection of various illnesses such as cancer and tuberculosis (Zhou, Greenspan & Shen, 2017: 179 - 192). Early detection of cancerous tumors increases a patient’s survival chances because proper treatment can be administered in a good time. Deep learning (DL) is an advancement of machine learning. It entails development and application of algorithms (neural nets) that try to simulate the working and functioning of the human brain (Chah, 2019). Unlike the traditional machine learning algorithm, the deep learning models require minimal data pre-processing steps. The models can process unstructured data such as images and text, and automatically extract hidden features and patterns from the dataset. The Convolution Neural Network (CNN) is the preferred deep learning model for computer vision tasks (Mishkin, Sergievskiy & Matas, 2017).

The exponential rise in big data, mainly because of the wide use of the internet, resulted in the development of state-of-art pre-trained deep learning models. The state of art models is usually trained on a very large dataset that can run to thousands or millions of data pieces. The patterns learnt by the models from the large training dataset can then be used to solve a similar or related task (Yang, Zhang, Tao & Ma, 2020). The aim of this study is to assess the performance of some of the state-of-art CNN models, in the automatic detection of COVID-19 and non-COVID-19 viral pneumonia. The dataset used to train, test and validate the Transfer learning models was made up of 4 main classes namely COVID-19, Viral pneumonia, Normal and Lung opacity. Lung opacity class refers to chest X-rays that indicate a patient has a lung infection that is neither COVID-19 nor viral pneumonia.

2. RELATED RESEARCH WORK

Krittanawong et al. (2019) discussed the application of deep learning techniques in cardiovascular medicine. Image segmentation can now be used for better analysis of cardiac imaging. Wearable devices such as smart watches can now be used to capture cardiovascular related data. The captured data can then be assessed by the deep learning models and the state of health a patient’s cardiovascular system successfully determined. The authors also suggested the automation of medical imaging analysis and interpretations as one of the major benefits of deep learning. The need for large amounts of data during model training, difficulty in applying DL techniques in clinical trials, black box decision making in model prediction, and unavailability of a standardized way when designing the deep learning models were indicated as some of the limitations of using deep learning in the field of medicine.

The present-day advancements in the field of Artificial Intelligence (AI) has facilitated the need for automating the interpretation of medical imaging. The analysis of chest X-rays has widely been proposed for automation because the chest X-ray machines are relatively inexpensive, chest X-rays are commonly used across the globe, and it takes several years to train a radiologist who can interpret the images (Kallianos et al., 2019). Dahmani et al. (2020) proposed a robust system that will help individuals with motor disabilities to move effortlessly using an electronic eye-controlled wheelchair. The authors developed and tested the system using four different techniques. The technique that used Convolution Neural Networks (CNNs) emerged as the most effective. The best performing CNN registered an accuracy of 99.3%. Hence, it was selected as the most appropriate model to be used in the gaze estimator which controls movements in the wheelchair.

Early detection and treatment of cancer is one of the safest ways to increase chances of survival for cancer patients. In the year 206 over 2.8 million women across the globe were diagnosed with cancer. Mammography is the technique that has been commonly used to diagnose breast cancer patients in the last 20 years. However, the technique has reported high false positive rates. The adverse side effects of mammography make it necessary to develop alternative Computer Aided Diagnostic (CAD) tools that facilitate a seamless and accurate detection of breast cancer (Mambou et al., 2018). Oh et al. (2020) developed a CNN architecture consisting of thirteen layers which was used to diagnose patients with Parkinson disease (PD). PD is caused by the constant decline in the brain motor function capability. The authors utilized electroencephalogram (EEG) signals from both healthy and PD patients as their model dataset. Their neural net performed well, generating accuracy, sensitivity, and specificity of 88.25%, 84.71% and 91.77% respectively.

Jo, Nho & Saykin (2019) reviewed multiple publications that proposed ways of using deep learning and computer vision to help in diagnosing patients with Alzheimer's disease (AD). The advancements in neuroimaging technologies has resulted in increased use of multi-channel neuroimaging dataset. Some of the reviewed research papers used a mixture of both the conventional machine learning and advanced deep learning methods, whereas others used a pure deep learning approach. A combination of both the machine learning and deep learning techniques generated an accuracy of 98.8% and 83.7 for the identification of AD and mild cognitive impairment conversion respectively. The pure deep learning methods produced accuracies of 96% and 84.2% for the identification of AD and mild conversion respectively.

Choi et al. (2021) developed a deep learning model using both long short-term memory (LSTM) and convolutional neural network (CNN), to help in stroke prediction. Stroke is predominant among the elderly. Stroke victims require constant observation because the disease has a high probability of recurrence, and also has a high death rate. The commonly used methods for stroke examination include using CT scans and Magnetic resonance imaging (MRI) scans. However, the two techniques require operating bulky and expensive machines, and it is also time-consuming. To help address some of the limitations associated with the existing approaches, the proposed technique uses raw electroencephalography (EEG) for model training and prediction. The developed model recorded an excellent stroke prediction accuracy of 94%.

Malaria is one of the leading causes of deaths especially in the tropics. The conventional way of diagnosing the illness is by testing blood smears of patients. If red blood cells infected by the malaria parasite can be identified in the smear by the help of a microscope, then a patient is determined to be malaria positive. This conventional testing method is inefficient and heavily relies on skills, expertise and experience of the person conducting the test. A new technique that uses deep transfer learning models has been proposed to automatically identify parasite infected red blood cells in the thin blood smears (Shekar, Revathy & Goud, 2020). More than a hundred respiratory illnesses share cough as a common symptom. Coughs have feature rich audio signals that can be leveraged to identify the cause of a respiratory problem. Kumar et al. (2020) came up with a deep learning model that can successfully detect ten different pulmonary diseases using cough signals.

Gao & Qian (2017) developed a CNN architecture that was used to detect and categorize five different strains of Tuberculosis (TB) in patients' CT images. Image segmentation was applied to the CT images converting them into rectangular shaped patches of different height and width sizes. The images were later standardized into a 30 x 30-pixel size before the model training phase began. 300 datasets were used to evaluate the model performance. The model generated 0.22 Kappa score and an accuracy of 41%, which was the highest accuracy value in the competition. Marques, Agarwal & de la Torre Díez (2020) suggested a novel way of working with a customized EfficientNet model architecture to automate the detection of COVID-19. A 10-fold stratified cross-validation was used to validate the results given by the CNN. Two major experiment types were conducted. The first one was binary classification that consisted of healthy and COVID-19 classes. The second one was multi-class classification that entailed healthy, COVID-19 and pneumonia classes. Chest X-ray images were used as the dataset for training, testing and validating the EfficientNet CNN. The binary and multi-class classification generated an accuracy of 99.6% and 96.7% respectively.

According to Rajaraman and Antani (2020), reverse transcription-polymerase chain reaction (RT-PCR) is a testing method frequently employed in the diagnosis of COVID-19. However, the method has several limitations such as it is labor intensive, it is expensive to set up its operating infrastructure, and it is also susceptible to giving false positive results. On the other hand, the use of computed tomography (CT) images is also prone to cross-contamination incidents. The authors proposed an approach of using weakly labeled dataset and data augmentation to recognize COVID-19 related opacities in the chest X-rays.

Depression is also one of the leading mental disorders that is adversely affecting people of different ages across the globe. Saidi, Othman & Saoud (2020) came up with a hybrid CNN and Support Vector Machine (SVM) machine that can automatically detect a patient suffering from depression. The hybrid model generated an accuracy score of 68% which was approximately 10% more than the accuracy results given by a pure CNN. Rahman et al. (2021) examined the effects of lung segmentation and image enhancement on the chest X-ray dataset used for COVID-19 detection. The authors analyzed different image enhancement methods such as gamma correction, image complement and histogram equalization (HE). Six pre-trained CNNs and a single shallow CNN model were utilized in the research experiment. The gamma correction technique gave the best performance metrics. However, when using the segmented dataset, the models generated a significantly superior output with an accuracy, precision, sensitivity, F1-score and specificity of 95.1%, 94.6%, 94.6%, 94.5 and 95.6% respectively

3. METHODOLOGY

The study used Transfer learning as the main technique for training, testing, and validating the deep learning algorithms in the identification of COVID-19 and viral pneumonia. A total of seven pre-trained CNNs were trained on a dataset of chest X-rays that was made up of 4 classes namely COVID-19, Viral pneumonia, Lung opacity and Normal. The performance of the CNN models was later evaluated using various metrics to determine which model was most effective at the classification task. In this section, Deep Transfer learning and other methods used to conduct the research have been discussed in detail.

3.1. Deep learning (DL)

Deep learning (DL) refers to a sub-branch of machine learning that utilizes neural networks with hierarchical architectures to learn hidden patterns and features from a dataset (Wooldridge, 2020). The neural networks try to imitate the functioning of the human brain. The advancement in computer processing power and the production of low cost computer hardware resources contributed to the rise of deep learning. Moreover, deep neural networks can effectively learn features from both non-linear and linear datasets, unlike the conventional machine learning models which are primarily applied on linear datasets (Mallick, Dhara & Rath, 2021). Deep learning techniques have been successfully utilized in a variety of fields such as natural language and image processing, voice detection and numerical data prediction. The major challenges associated with DL are the need for a large training dataset and also in most cases training DL models is time-consuming and requires computers with a significantly powerful processing capability (Wooldridge, 2020).

Despite being a sub-branch of machine learning (ML), deep learning differentiates itself from the traditional ML in two main ways. The first one is the nature of the dataset used to train and test the deep neural networks. The conventional ML algorithms primarily use structured, well labeled datasets to learn and make predictions (Kaur, 2022). In situations where the dataset is unstructured, pre-processing must first be applied to the dataset to convert it into a structured format. Deep learning notably minimizes the many data pre-processing steps associated with ML. The deep neural networks can automatically extract features and other hidden patterns from unstructured data like images and text. The second difference between ML and DL is in the size of the training dataset. DL models typically require a larger dataset to successfully train and learn meaningful information from the data. Moreover, the DL models also take a longer time to train unlike most traditional ML models which take a shorter time to train and may require a smaller size of training data (Kaur, 2022).

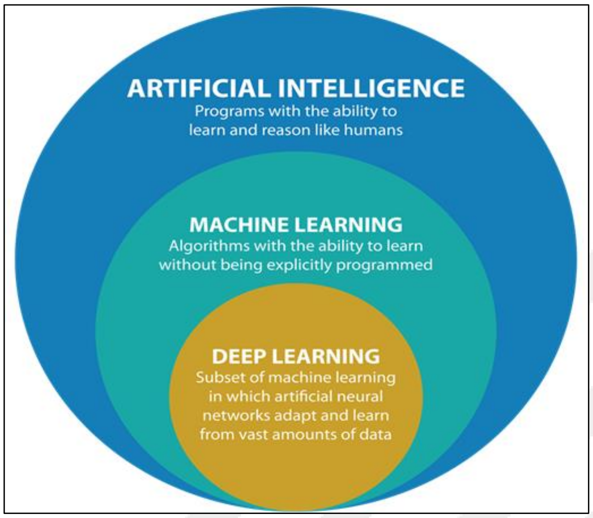


Figure 3. 1. Deep learning as a sub-branch of Machine Learning

A connection of several layers which contain multiple interconnected nodes make up the deep neural networks architecture. The layers forming the neural nets are organized in a hierarchical structure, that is, one layer lying on top of the other (Michelucci, 2018). The main task of each layer in the architecture is to refine the features learned by the preceding layer. The input and classification layers are situated at the extreme opposite ends of the model structure, and are known as visible layers (Karoly, Galambos, Kuti & Rudas, 2021). The input layer is responsible for taking in the data for model processing, whereas the classification layer is responsible for categorizing data into the various dataset classes. Both forward and backward propagation are the two key strategies used by the deep learning models to learn hidden patterns from data, make model predictions and adjust any encountered errors made by models during prediction (Lee, Park, Kim, Yu, Jo, & Choi, 2019). Gradient descent is one of the algorithm commonly used by the neural nets to compute the error margin generated during model prediction. Once the error size is determined the weights and biases of the models are adjusted accordingly by moving backwards across all existing layers in their structure (Karoly et al., 2021).

3.2. Convolution neural network (CNN)

CNN is the most popularly used neural net for computer vision tasks. Three key layers make up a CNN architecture namely convolution layers, fully connected or dense layers, and pooling layers (Zhang, Tian & Li, 2018). Each of the three layers have their own specific function. A CNN is trained in two phases. The first one known as forward phase, entails representing the input data using the existing layer parameters. A prediction is then made by the output (classification) layer and the prediction error size is determined by making comparison with the ground truth labels. The second stage, known as backward phase, entails calculating the gradient of each layer parameter and adjusting the weights and biases accordingly (Lee et al., 2019). Figure 3.2 shows the structure of CNN

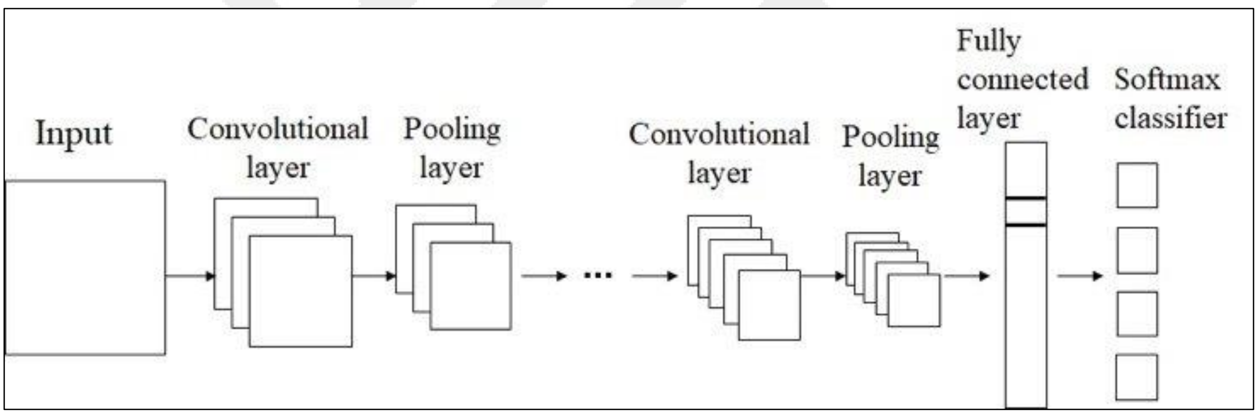


Figure 3. 2. General structure of a Convolution Neural Network (CNN)

3.2.1. Types of CNN layers

In this section, three main types of CNN layers that made up the deep neural network models used in the study have been discussed. The general structure and functioning of the layers have been elaborated in detail.

<u>Convolution layers</u>

The primary building block of any CNN architecture is the convolution layer. The layer contains kernel (filter) parameters which performs a convolution operation on the input image (Zhang et al., 2018). The convolution operation entails sliding the filter across the width and height of the input data. As the filter is sliding across the input data, a dot product operation is done between the filter elements and the input. The learning process of the convolution layers is also faster than the conventional dense layers. The speedy learning is one of the key reasons why they are preferred over the fully connected layers for the computer vision tasks. Figure 3.3 Shows the convolution operation.

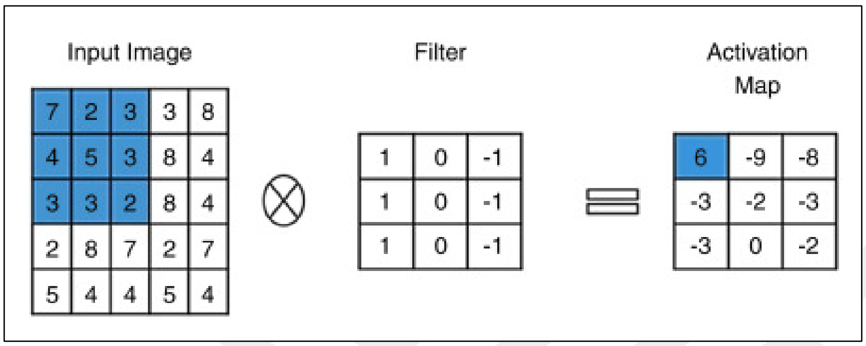


Figure 3. 3. The convolution operation

<u>Pooling layers</u>

The pooling layers are usually stacked after the convolution layers in a CNN architecture. Their primary role in the architecture is to shrink the size of feature maps (Lee, Gallagher & Tu, 2018). The most prominent pooling strategies are Max and Average pooling. In Max pooling operation, the filter picks the greatest member in the feature map area it covers. Hence, the max pooling layers generate output feature maps that have the most significant features of the input feature maps. Figure 3.4. shows max pooling operation in action. In average pooling operation, the filter calculates the mean of all the elements it covers in a given feature map area. Hence, the average pooling layers will produce output feature maps containing the average of input feature maps members. Figure 3.5 shows the average pooling operation.

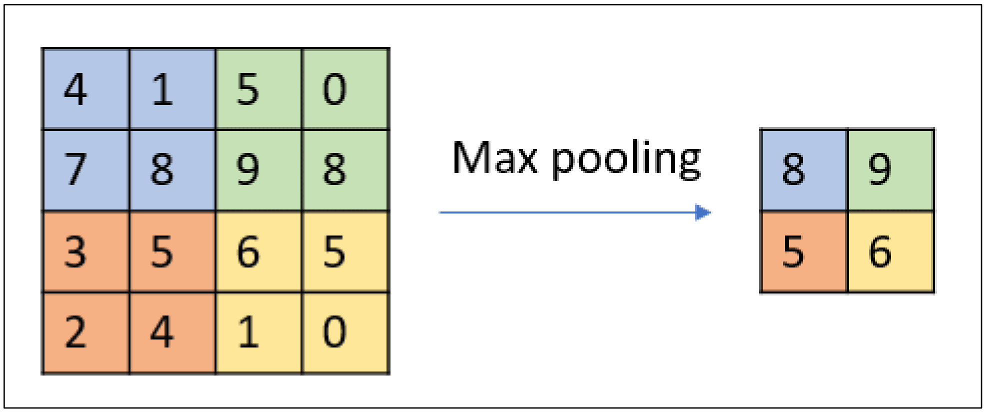


Figure 3.4. Max pooling operation

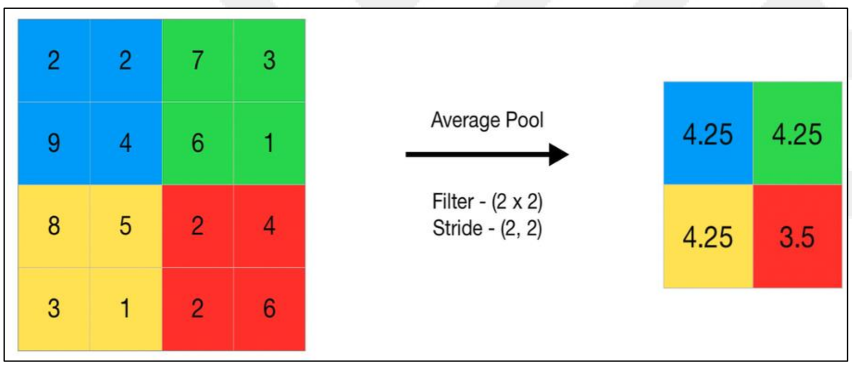


Figure 3. 5. Average pooling operation

<u>Fully connected layers</u>

The fully connected layers (also referred to as dense layers) are typically stacked at the tail end of the CNN architecture. A fully connected layer in a CNN obtains its input from the last convolution or pooling layer. The input to the fully connected layer is first flattened into a one-dimensional feature vector and later directed to the layer (Basha, Dubey, Pulabaigari & Mukherjee, 2020). Figure 3.6 showing flattening operation. The fully connected layers can either be single or multiple depending on the nature of task to be executed. They also perform similar mathematical operations as the conventional Artificial Neural Networks (ANN). The very last layer in the model architecture, also known as the out-put or classification layer (for a classification task), uses various activation functions such as softmax and sigmoid to get the prediction probabilities of the model (Basha, 2020). The fully connected layers also contribute the greatest number of all parameters found in a CNN. The high number of parameters makes them require a lot of computational power for their successful training.

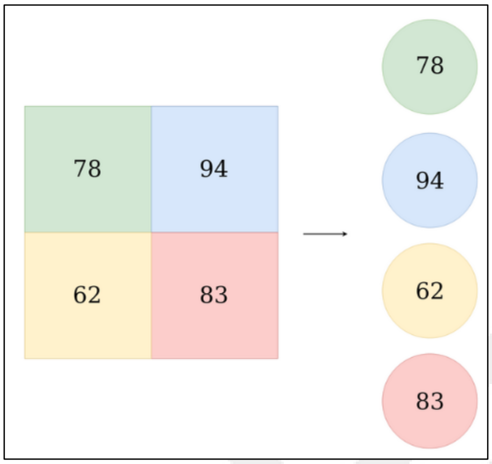


Figure 3.6. Flattening operation

3.3. Model training strategies

Unlike the conventional machine learning models and other shallow architectures, deep learning models are very good at identifying hidden patterns and features in large datasets. Nevertheless, the deep architecture means the models have a large number of parameters which makes them susceptible to overfitting. In recent years, several regularization techniques have been developed to help alleviate the overfitting challenge. Some of these methods are discussed below.

3.3.1. Dropout

Dropout protects a deep learning model from overfitting by arbitrarily inactivating nodes and their associated paths in the neural network (Hahn & Choi, 2020). The random node inactivation ensures the deep neural net does not depend on a specific set of neurons in the network. It impels all nodes present in the network to learn concealed patterns and features from the training dataset enabling them to generalize well. Figure 3.7 demonstrates how Dropout works.

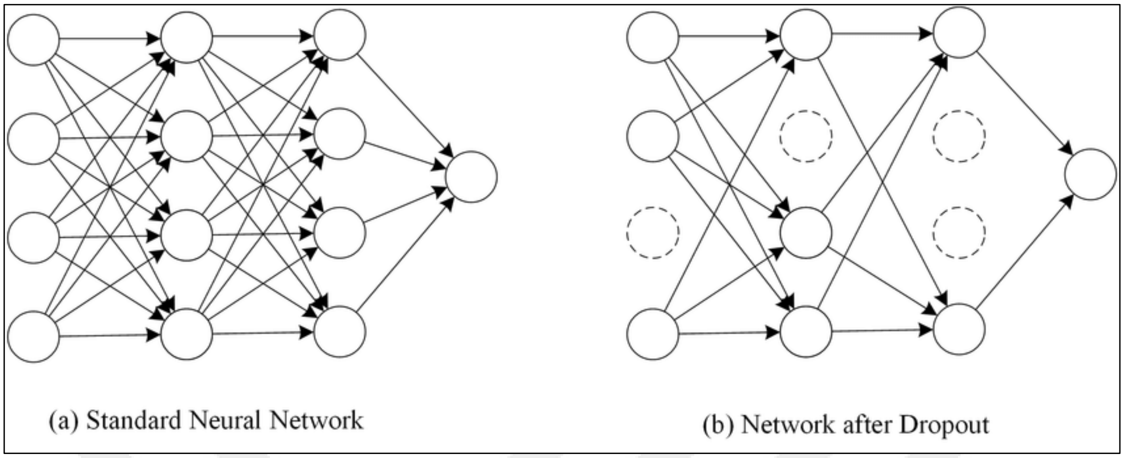


Figure 3. 7. Dropout operation in a neural network architecture

3.3.2. Data augmentation

Data augmentation can be defined as the process of introducing heterogeneity in a dataset by making minor changes on copies of the current data (Lemley, Bazrafkan & Corcoran, 2017). Various transformation methods can be used to generate more data using augmentation such as zooming in and out, image cropping, and image rotation. The transformation techniques can also be applied simultaneously to further increase the number of distinct training dataset. Apart from geometrical manipulation, other effects such as modifying color properties, adding filters and noise can also be used as an augmentation technique especially when dealing with image datasets (Basha, 2020). Figure 3.8 gives an example of how three cat images are augmented to produce other cat images with different properties. Data augmentation is mainly utilized in supervised learning because of the presence of already tagged data. This helps us to elude the additional cost of labeling the new data generated by augmentation.

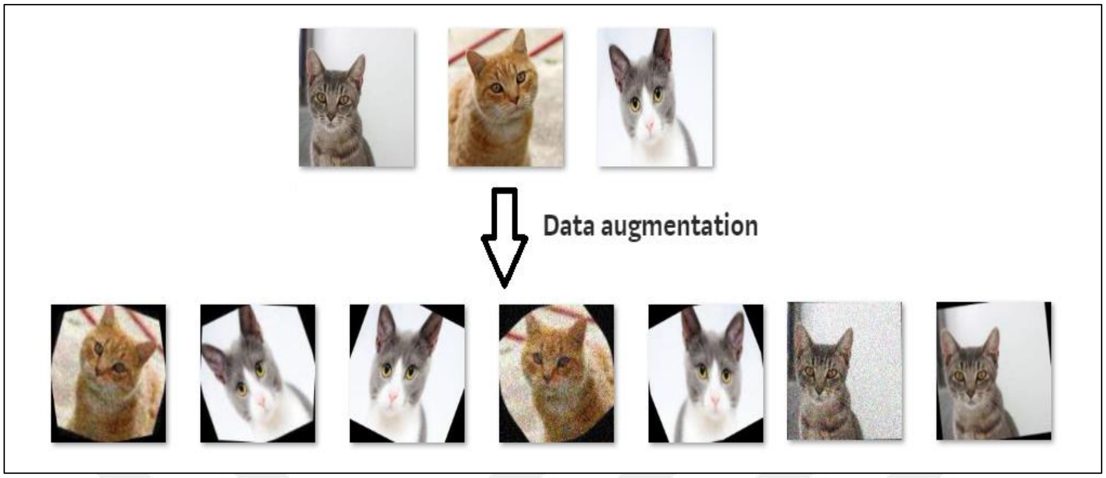


Figure 3. 8. Image augmentation in action

3.4. Transfer learning

Gathering enough data when dealing with a novel task may prove to be a difficult task. Moreover, generating excellent model results using a small sized dataset can be very challenging or in some cases it can be unattainable. However, transfer learning helps us to address this limitation by allowing us to use a small dataset to train our models but still obtain excellent model metrics (Lu, Behbood, Hao, Zuo, Xue & Zhang, 2015). Transfer learning can be defined as the process of reusing a pre-trained model as a basis for training another model on a new but related task (Day & Khoshgoftaar, 2017). In transfer learning, sharing of generalized knowledge between the pre-trained model and the newly created model is possible because the models are meant to perform analogous tasks. Transfer learning helps us to minimize the resources required to build and train our models such as time and size of dataset. Figure 3.9 illustrates a high level overview of the operation of Transfer learning. Unlike constructing models from scratch, transfer learning models require less amount of time to build and utilize less data in training. However, the models are still able to generate better results than their counterparts created from scratch (Lu et al., 2015).

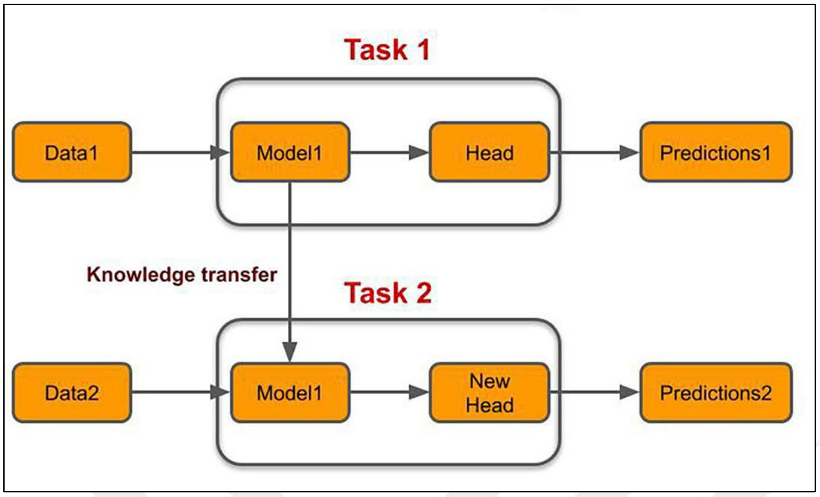


Figure 3. 9. Transfer learning in action

3.4.1. Common transfer learning techniques

Figure 3.10 shows the various categories of Transfer learning techniques. Data availability, application domain and the nature of tasks are the three main factors that determine the transfer learning technique to be utilized (Niu, Liu, Wang & Song, 2020). Conventionally, transfer learning has been classified into three namely Inductive, Transductive and Unsupervised Transfer learning (Day & Khoshgoftaar, 2017). This classification type is based on the number of labeled and unlabeled dataset, and the nature of domain task. In inductive transfer learning, both the target domain and source domain are identical. However, the nature of the task the model is executing at the target and the source is divergent. Knowledge from the pre-trained model (also known as source model) is used as the basis by the new model to enhance the target task.

Self-taught and multi-task learning are other subdivisions that can be obtained from inductive transfer learning depending on whether data contained in the source domain is labeled or unlabeled (Niu et al., 2020). In unsupervised transfer learning, the dataset contained in target task and the source task are unlabeled. However, all the other characteristics of this transfer learning technique are identical to inductive transfer learning. In transductive transfer learning, the target domain and the source domain lack an exact match. Nevertheless, the target task and source task have some resemblance. In most cases, the target domain of this transfer learning technique contains unlabeled data, whereas the source domain contains multiple labeled data.

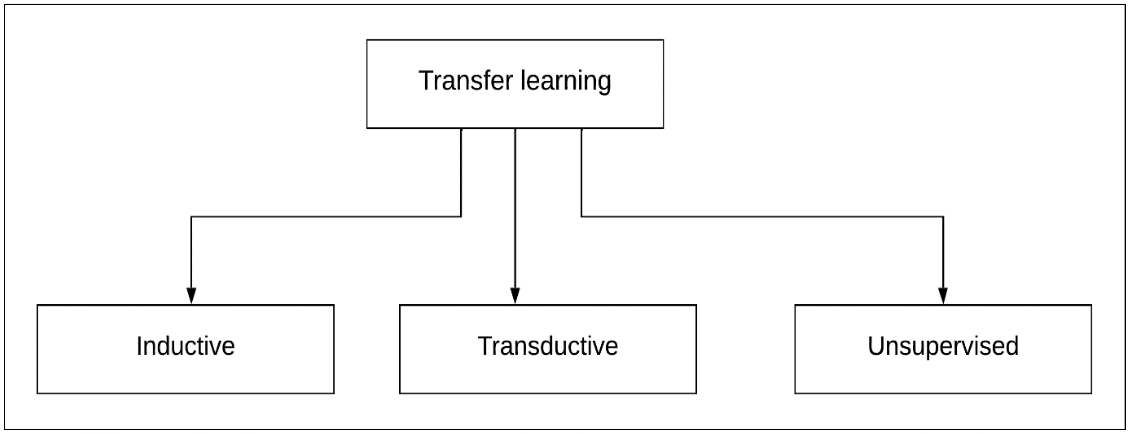


Figure 3. 10. Commonly used transfer learning techniques

3.4.2. Transfer learning techniques for deep learning

Over the recent years, several state-of-the-art deep learning systems have been developed to help solve complex problems across various domains and industries. In most cases, research teams and individuals responsible for building the state-of-the-art deep neural nets share information about their neural nets with the public, to support research in the field of AI. These pre-trained neural nets are the back-bone of transfer learning with regards to deep learning. The application of transfer learning in deep learning can also be referred to as deep transfer learning (Zhang, Zhao & Wu, 2020). In this section, we will discuss two of the commonly utilized deep transfer learning strategies which are Feature extraction and Fine-tuning transfer learning. In this paper, both Feature extraction and Fine-tuning were used to train the CNN models. Figure 3.11 shows the general steps followed during the model training process. The neural nets were trained using three main strategies. The first one was using the pre-trained CNNs as a feature extractor and a fully connected dense layer as the final classification layer. The second one was using the pre-trained CNNs as a feature extractor and a Support Vector Machine (SVM) as the classifier. The third one was fine-tuning the pre-trained CNNs.

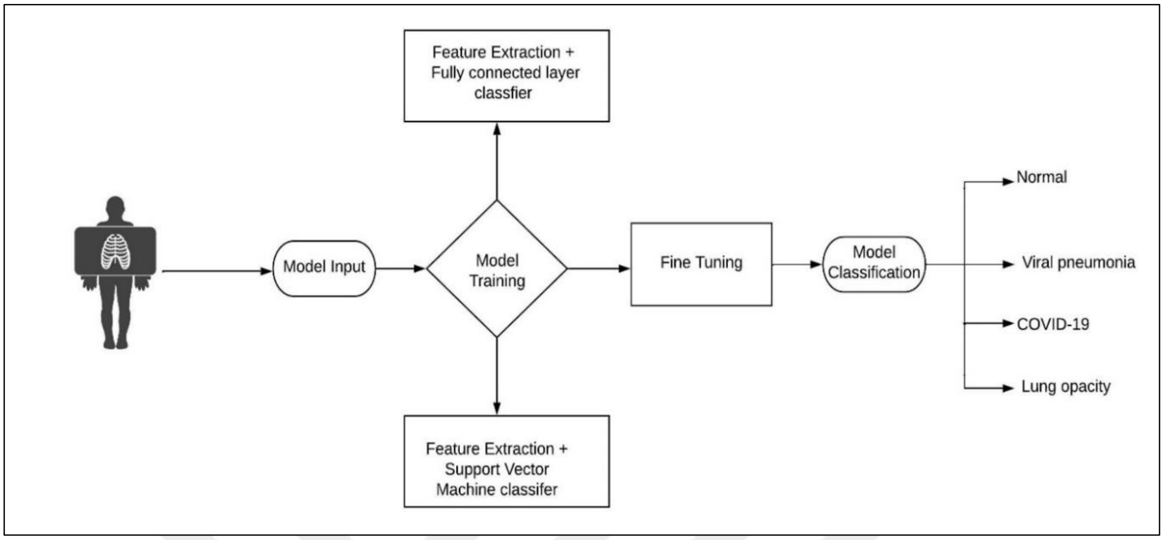


Figure 3. 11. Steps followed during the model training process

<u>Ready-made pre-trained models as feature extractors</u>

Deep neural nets are composed of a layered architecture. Different layers learn different features from the input data. The deeper layers in the architecture are responsible for identifying high level generalized features, whereas the shallower layers are responsible for learning features that are more specific to the dataset (Wiatowski & Bolcskei, 2018). The final layer also known as the output or classification layer (for classification tasks) is tasked with categorizing data into various classes the model is configured to classify. In supervised learning, the output layer is usually a fully connected dense layer. However, in some cases a conventional machine learning model such as SVM can also be used as a classifier (Wooldridge, 2020). This research work only focuses on classification tasks hence the output or final layer can be interchangeably referred to as classification layer. Figure 3.12 gives a general overview of how feature extraction by using an off-the-shelf-pretrained model is carried out.

The primary advantage of this strategy is the ability to utilize weights already learned by the pre-trained models to extract features from the new data (Wiatowski & Bolcskei, 2018). The weights of the pre-trained model are also not adjusted during training. The base layers of the pre-trained model are all frozen during training and this is why their weights are not updated. The original classification layer that came with the pre-trained model is cut off from the model and replaced with a single or multiple layers including a classification layer. The weights of newly fitted layer(s) are adjustable during training (Zhang et al., 2020).

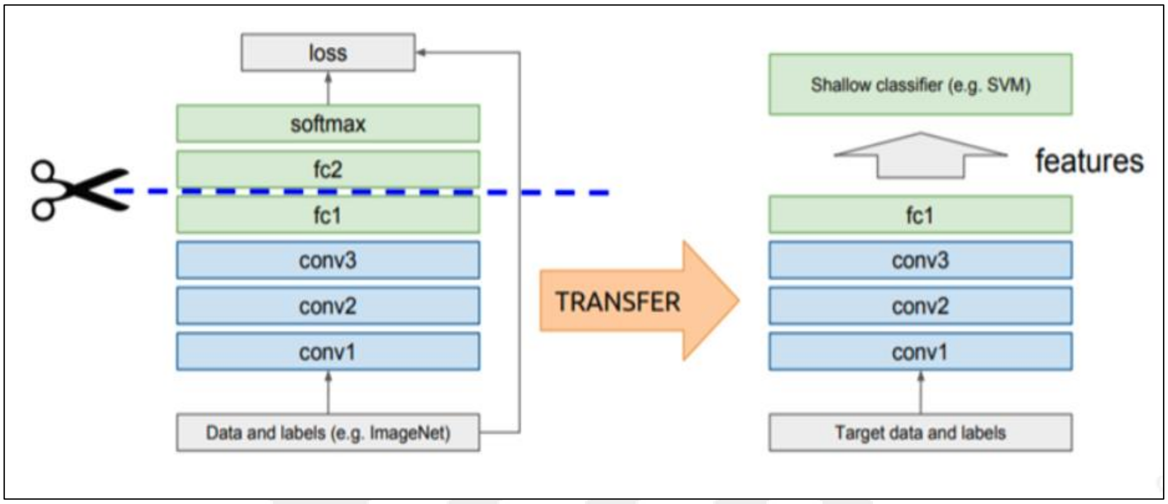


Figure 3. 12. Feature extraction by using an off-the-shelf pre-trained model

Fine-tuning ready-made pre-trained models

Unlike using the previously discussed technique, not only do we replace the pre-trained model output-layer but also select some of the model's top most layers for retraining (Wang et al., 2018). The layers to be unfrozen for training are identified through experimentation. In rare cases, the entire model layers can also be unfrozen for training. However, this may prevent us from realizing benefits of transfer learning such as faster training speeds and the model may also be prone to overfitting (Mutasa, Sun & Ha, 2020). In fine-tuning, the number of layers to be unfrozen can be additionally used as a hyper-parameter to enhance the overall model performance. Figure 3.13 illustrates the general model fine-tuning process.

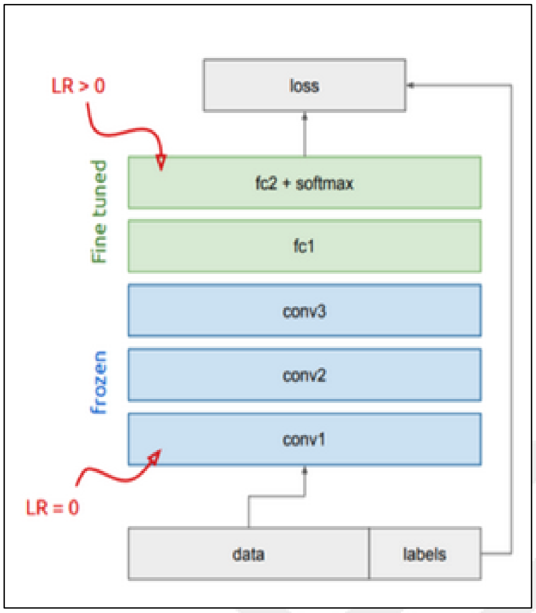


Figure 3. 13. Model fine-tuning

3.4.3. Stages in transfer learning

Transfer learning entails six main phases. The first phase is the selection of a ready-made pre-trained model. There are several already developed models suited for different tasks. Hence, our model choice should be guided by the nature of tasks we want the model to perform. There should be a strong interdependence between the target domain and the knowledge possessed by the pre-trained model, for an effective execution of transfer learning (Peirelinck et al., 2022). The second phase involves construction of a base model. At this stage, we can decide to train our model from scratch by using a bare neural net architecture or download the pre-trained model network weights that will help us drastically cut down our model training time. In the third phase, before training of the pre-trained neural net commences the model's underlying layers will first be frozen. Freezing the layers will ensure the pre-trained model will not lose all its learning (Guérin, Thiery, Nyiri, Gibaru & Boots, 2021). A model that loses its learning will have to be trained from scratch and this will require more time and computational resources.

In the fourth phase, new trainable layer(s) are added on top of the frozen base layers. We can decide to add a single output layer or several trainable layers depending on the nature of task we want to execute (Peirelinck et al., 2022). The trainable layers enable the neural net to learn finer-grain patterns associated with our input data. In a situation where the final output layer is a fully connected dense layer used for classification tasks. The number of nodes contained by the output dense layer should coincide with the number of classes in our dataset (Guo et al., 2016). For example, if the neural net is trained to classify data into four classes then the output fully connected dense layer should also have four nodes. Figure 3.14 shows an overview of how step four is implemented. The fifth stage involves the actual model training. The generated model results can further be enhanced through a series of experimentation and hyper-parameter tuning (Mutasa, Sun & Ha, 2020). Examples of the hyper-parameters include the learning rate, batch size, number of epochs, momentum, regularization constant etc. The sixth step is usually optional and is also dictated by the function our model will execute. The step entails fine-tuning our neural net to further enhance its performance. Small learning rate values are used during fine-tuning to protect the models from overfitting.

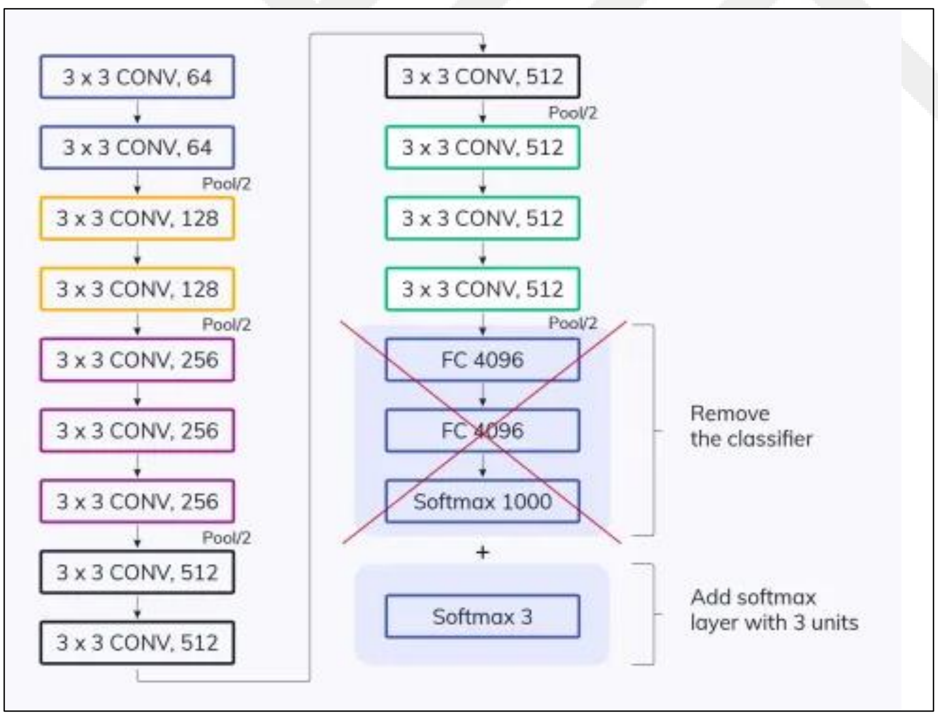


Figure 3. 14. Creating a base model with a custom output layer

3.5. Dataset used in the study

The chest X-ray images used to train the CNN models in the study were obtained from an open-access database created by a group of researchers in conjunction with medical doctors from different parts of the world (Chowdhury et al., 2020). The database was set up by gathering chest X-ray images from various sources such as publicly accessible github repositories, different research publications and the Italian Society of Medical and Interventional Radiology (SIRM). The dataset is hosted in the Kaggle website under the name COVID-19 Radiography Database (Kaggle, 2022). As at the time of collection, the dataset was made up of 10192 normal, 1345 viral pneumonia, 3616 COVID-19, and 6012 lung opacity images. Table 3.1. gives a summary of the number of X-ray images used in the study.

Table 3. 1. Number X-ray images used for training, validation and testing in each class

Classes	Total Number of X-ray Images	Training set (Before Augmentation)	Training set (After Augmentation)	Validation set	Testing set
Normal	10192	7134	7134	1019	2039
Viral Pneumonia	1345	941	7110	135	269
COVID-19	3616	2531	7035	361	724
Lung Opacity	6012	4208	7120	601	1203

3.6. Model selection

In this study, seven different pre-trained CNN models were trained, validated and tested. The models included MobileNetSmall, MobileNetLarge, EfficieNetV2B2, ResNet50V2, DenseNet 121, NasNetMobile and Xception. Two different versions of the MobileNetV3 were used to investigate how depth influences the performance of a model with similar structure. The Adam AMSGrad variant was used as the optimization algorithm. AMSgrad is a stochastic optimization technique that attempts to boost the convergence capabilities of the Adam algorithm, by limiting the occurrence of large instantaneous changes in the input variables' learning rates (Tran & Phong, 2019).

A constant learning rate of 0.001 was used when training the models using Feature extraction transfer learning with both a fully connected dense layer classifier and a SVM classifier. A slightly lower learning rate of 0.0001 was used when training the model using the Fine-tuning transfer learning to enable the CNN models to converge optimally and also avoid overfitting the models. Moreover, a dropout layer with a rate of 0.4 was also used in both feature extractions transfer learning with a dense layer as a classifier and fine-tuning transfer learning as a model overfitting deterrent. When training the CNNs using feature extraction with a SVM classifier, the dropout layer was not used and instead a L2 regularizer with a regularization factor of 0.0001 was utilized to protect the models against overfitting. All the hyper-parameter values used during the training process were determined after several experiments.

A global average pooling layer was also added to the architecture. The function of the layer is to compute the mean of each feature map and output a vector that is forwarded to the final output layer for classification. In this paper, the global average pooling layers was preferred over the traditional fully connected dense layers because they are naturally compatible with the convolution structure, and also they do not require any parameter optimization which may be resourceful in circumventing overfitting. Figure 3.15. shows the MobileNetV3Large architecture when using a fully connected dense layer as classifier for both feature extraction and fine-tuning transfer learning. Figure 3.16. illustrates MobileNetV3Large architecture when using an SVM classifier. A sigmoid activation function and binary_crossentropy loss function were used in binary classification, when using a fully connected dense layer as the classification layer. For the multi-class classification, the dense classifiers used categorical_crossentropy loss function and the softmax activation function. When training the models using feature extraction and a SVM classifier technique. The final output layer was converted to a SVM by using an L2 kernel regularizer and a hinge loss function and linear activation function for binary classification. For multi-class classification, a Squared_hinge loss function and softmax activation function were applied.

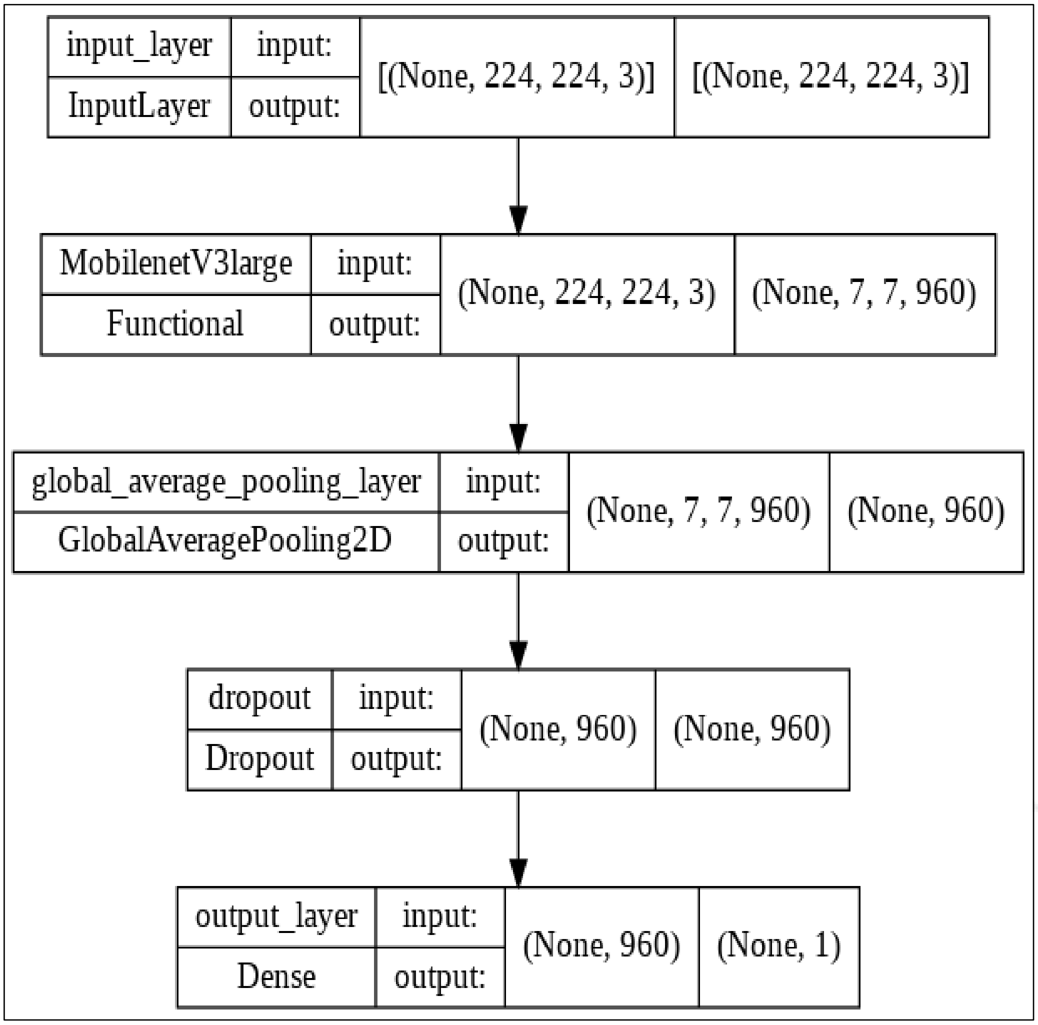


Figure 3. 15. MobileNetV3Large architecture when using a fully connected dense layer as classifier

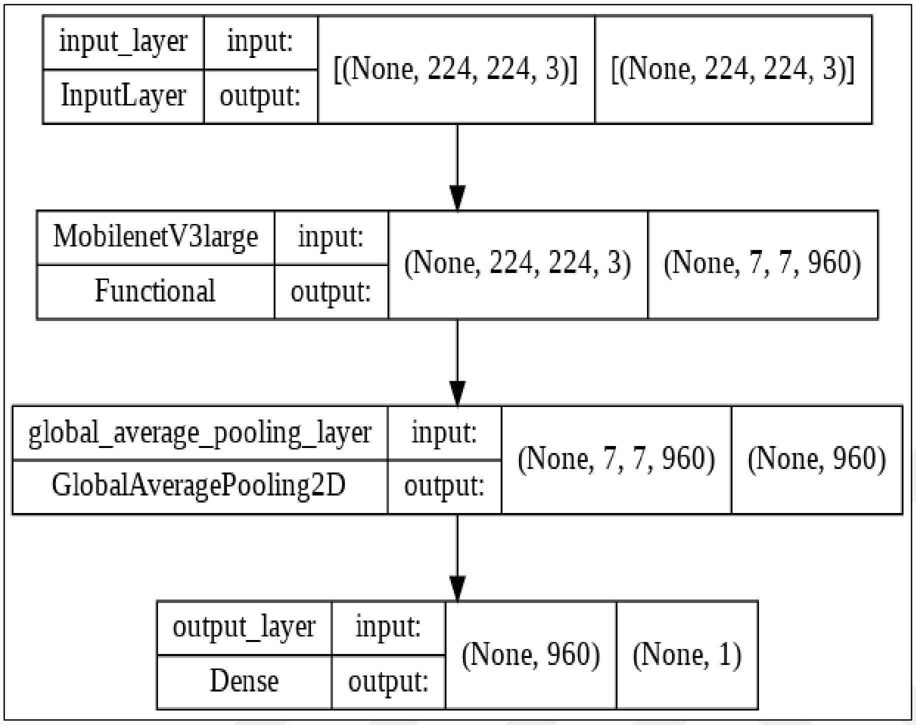


Figure 3. 16. MobileNetV3Large architecture when using an SVM classifier

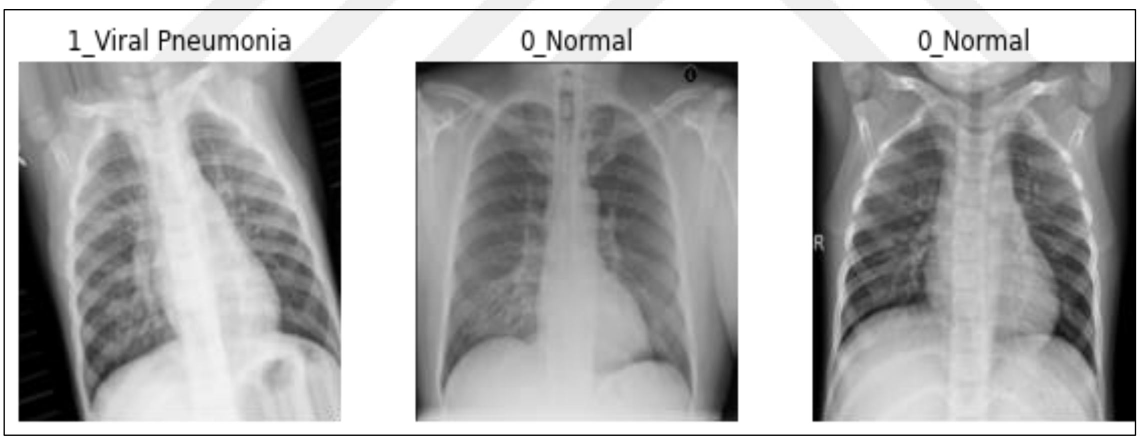
3.7. Data-preprocessing

The chest X-ray images were resized depending on the CNN model requirements. Most of the models required an image size of 224 x 224 pixels, other used image sizes were 260 x 260 pixels and 229 x 229 pixels. Table 3.2. shows a summary of the image size requirements by each of the models used in this research work.

Table 3. 2. Summary of the image size requirements

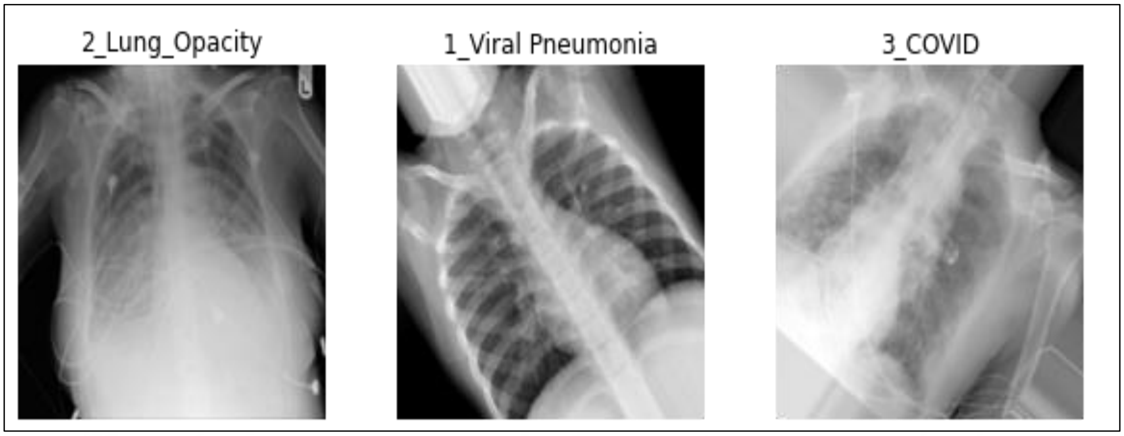
Model Name	Image size (Pixels)
EfficientNetV2B2	260 x 260
MobileNetV3Small	224 x 224
MobileNetV3Large	224 x 224
ResNet50V2	224 x 224
DenseNet121	224 x 224
NASNetMobile	224 x 224
Xception	229 x 229

Rotation based image augmentation technique was used to generate more training dataset. The additionally generated images ensured the number of images across the 4 classes used to train the models was balanced. A balanced training dataset eliminates any form of bias during model prediction phase, and enables the models to generalize well when given a new dataset. The X-ray images were first augmented and shuffled before being fed to the model for training. The normal chest X-ray images were the only training dataset not augmented, because the normal class contained the greatest number of X-ray images in the training dataset. The augmented images were rotated with an angle of between -40 to 40 degrees. Table 3.1 displays the number of X-ray images used to train, validate and test the pre-trained models. Figure 3.17. and 3.18. shows a sample of the chest X-ray images before and after augmentation respectively. Image augmentation was not applied to the testing and validation dataset. Random sampling was used to split the X-ray image dataset into 70% training, 20% testing and 10% validation data. A constant batch size of 32 images was also used during model training.



(a) (b) (c)

Figure 3. 17. Chest X-ray images before data augmentation



(a) (b) (c)

Figure 3. 18. Chest X-ray images after data augmentation

3.8. Performance metrics

Various metrics were used to evaluate the performance of the trained CNN models as follows: (a) True positive (TP), which are X-ray images correctly categorized as disease positive, (b) True Negative (TN), which are X-ray images correctly identified as disease negative or healthy, (c) False Positive (FP), which are X-ray images wrongly identified as a diseased case but are in fact a healthy case, (d) False Negative, which are X-ray images wrongly categorized as a healthy case but are in fact a positive case. When a trained model predicts a COVID-19 positive X-ray image as positive, the prediction is a TP. When the COVID-19 positive X-ray image is predicted as normal or healthy, the prediction is a FN. When a healthy X-ray image is categorized as COVID-19 positive, the prediction is FP. When a healthy X-ray image is predicted as normal, the prediction is TN.

The metrics discussed above in this section were used to compute the accuracy, recall and precision and F1-score of the CNN models. The equations 1, 2, 3 and 4 show how the metrics can be calculated.

$$\text{Accuracy} = TP + TN / (TP + TN + FP + FN) \quad (1)$$

$$\text{Recall} = TP / (TP + FN) \quad (2)$$

Precision = TP / (TP+FP) (3)

F-score = 2 * (Precision * Recall) / (Precision + Recall) (4)

A high recall value means the number of False Negatives will be low. Low False Negative values ensures the number of COVID-19 or Viral pneumonia positive patients incorrectly diagnosed as healthy is minimized. Accurately diagnosing COVID-19 patients helps to eliminate the risk of the patients unknowingly spreading the illness to other healthy people. A high precision means the number of False Positives will be low. Low False Positive values guarantee the number of healthy individuals incorrectly diagnosed with either COVID-19 or viral pneumonia is minimized. This will ensure resources dedicated for the treatment and limiting the spread of the diseases is administered appropriately.

4. FINDINGS AND DISCUSSION

In this study, two major types of categorizations were evaluated. The first one was binary classification, which consisted of normal vs viral pneumonia and normal vs COVID-19 classes. The second one was multiclass classification which consisted of normal vs viral pneumonia vs COVID-19, and normal vs viral pneumonia vs COVID-19 vs lung opacity. The lung opacity class refers to other lung infections that are neither COVID-19 nor viral pneumonia. The models were also trained using three varying deep transfer learning strategies discussed in the methodology section. The first strategy was using the ready-made pre-trained neural nets as feature extractors and a fully connected dense layer as the final classification layer. The second strategy is similar to the first one with the only exception being a support vector machine was used as the final output layer for the neural nets. The third technique entailed fine-tuning the pre-trained neural nets with an aim of enhancing their performance.

4.1. Pre-trained models as feature extractors and fully connected layers as classifier

In this model training and classification category, all the underlaying layers with exclusion of the top most classification layer of the pretratrained models were frozen. This top most output layer was then replaced by another fully connected layer that contained the number of nodes corresponding to the number of classes the model was going to classify.

4.1.1. Binary classification: normal vs COVID-19

Table 4.1. illustrates the accuracy, precision, recall, and F1-score values of the pre-trained models in classification of normal and COVID-19 chest X-ray images. The results generated by the models were promising. The pre-trained CNNs generated accuracy values ≥ 91%, precision values ≥ 91%, recall values ≥ 86% and F1-scores ≥ 88%. The EfficientNetV2B2 neural net performed best with an accuracy, precision, recall and F1-Score of 99%, 98%, 98%, 98% respectively. Figure 4.1. shows the confusion matrix of the EfficientNetV2B2 model. Out of the 724 COVID-19 X-ray images 646 were correctly categorized as COVID-19 positive and 78 were incorrectly categorized as normal. Out 2039 normal X-ray images 2013 were correctly identified as normal and only 26 were wrongly classified as COVID-19 positive. Figure 4.2. shows the time taken to train the pre-trained CNNs in the normal vs COVID-19 classification. Figure 4.3. presents the training accuracy and loss curves of the EfficientNetV2B2 in this

classification category. MobileNetV3Small and MobileNetV3Large recorded the best and an almost identical training time of 200 seconds and 204 seconds respectively. ResNet50V2 and DesnseNet121 also had comparable training time of 491 seconds and 492 seconds respectively, a difference of only a second. Xception neural net had the worst training time of 714 seconds.

Table 4. 1. Performance of the pre-trained neural nets in the classification of normal and COVID-19 chest X-rays

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	99	98	98	98
MobileNetV3Small	99	96	97	97
MobileNetV3Large	99	97	98	97
ResNet50V2	94	94	91	92
DenseNet121	91	92	86	88
NASNetMobile	92	91	88	89
Xception	93	93	90	91

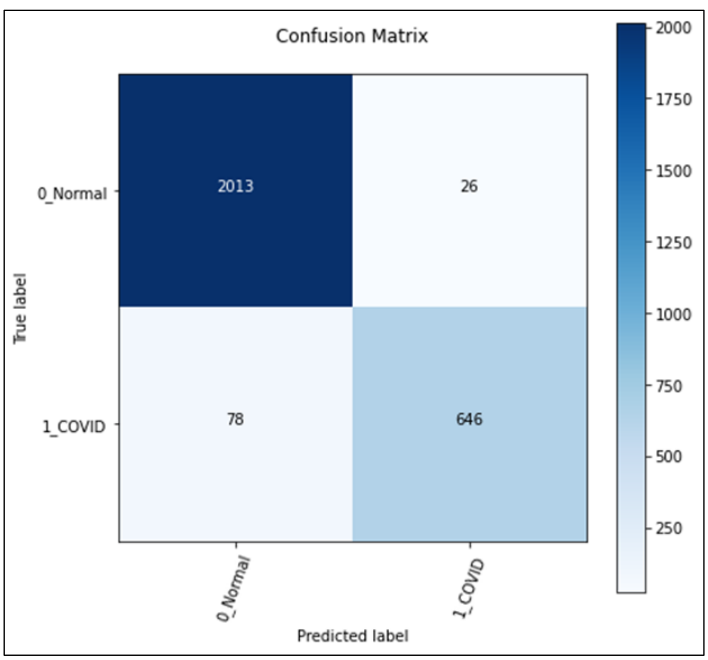


Figure 4. 1. Confusion matrix of EfficientNetV2B2 in the classification of normal and COVID-19 X-ray images.

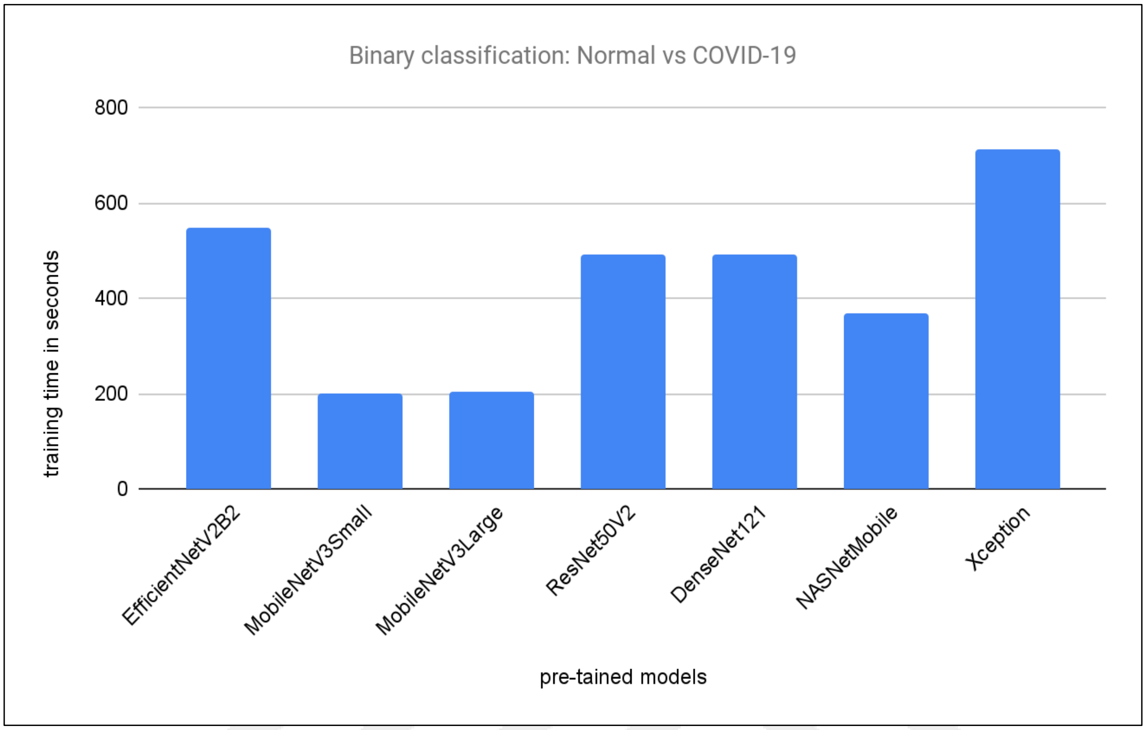


Figure 4. 2. Normal vs COVID-19 model training time.

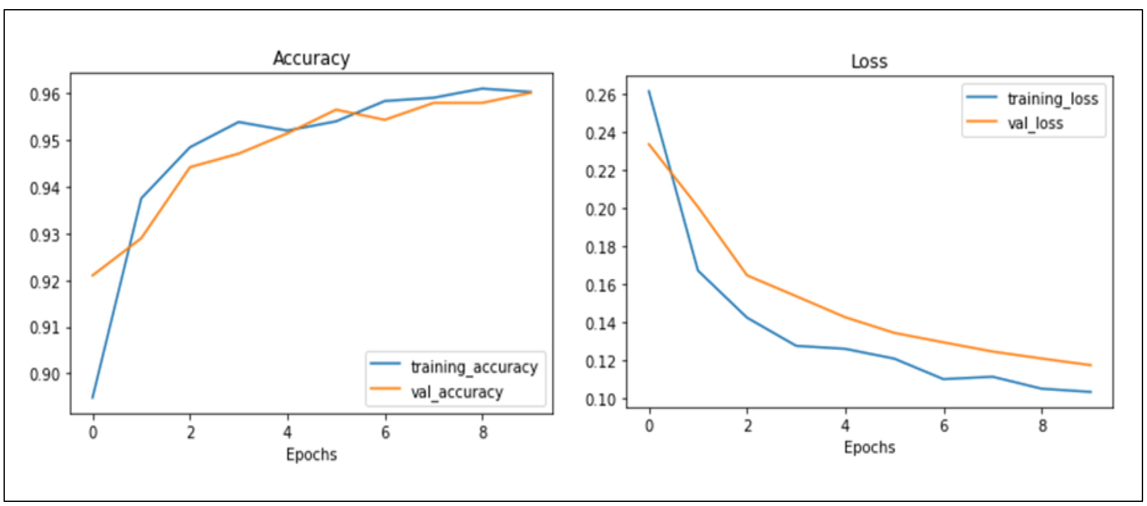


Figure 4. 3. EfficientNetV2B2 accuracy and loss curves

4.1.2. Binary classification: normal vs viral pneumonia

Table 4.2. illustrates results generated by the pre-trained CNNs in the identification of normal and viral pneumonia chest X-rays. All the CNNs generated accuracy scores ≥ 97%, precision

scores ≥ 95%, recall scores ≥ 92% and F1-Scores ≥ 93%. The EffcientNetV2B2 was the best performing CNN with an accuracy, precision, recall and F1-Score of 99%, 98%, 98%, 98% respectively. The NASMobile recorded the lowest performance values in all the four metrics with an accuracy of 97%, precision of 95%, recall of 92%, and F1-Score of 93%. Figure 4.4. shows the confusion matrix of the best performing CNN. Out of 269 viral pneumonia images, 257were correctly categorized as viral pneumonia positive and only 12 were misclassified as normal X-ray images. Out of 2039 chest X-ray images, only 8 were erroneously categorized as viral pneumonia positive. Figure 4.5. displays the time taken to train the pre-trained CNNs in the classification of normal and viral pneumonia chest X-rays. Figure 4.6. illustrates the accuracy and loss curves when training the EfficientNetV2B2 neural net in this classification category. MobileNetV3Small had the shortest training time of 190 seconds followed closely by MobileNetV3Large which recorded a training time of 197 seconds. Xception had the longest training time of 716 seconds. Despite the remarkable model performance metrics EfficientNetV2B2 had the second longest training time of 548 seconds.

Table 4. 2.The accuracy, precision, recall, and F1-score values of pre-trained CNNs

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	99	98	98	98
MobileNetV3Small	99	96	97	97
MobileNetV3Large	99	97	98	97
ResNet50V2	99	98	96	97
DenseNet121	98	96	97	96
NASNetMobile	97	95	92	93
Xception	99	97	96	96

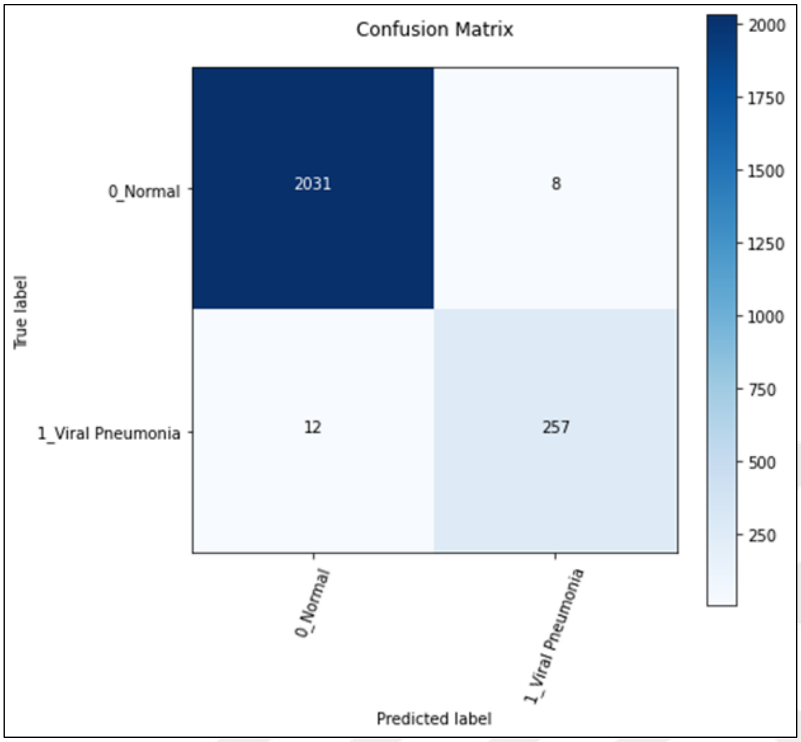


Figure 4. 4. Confusion matrix of EfficientNetV2B2

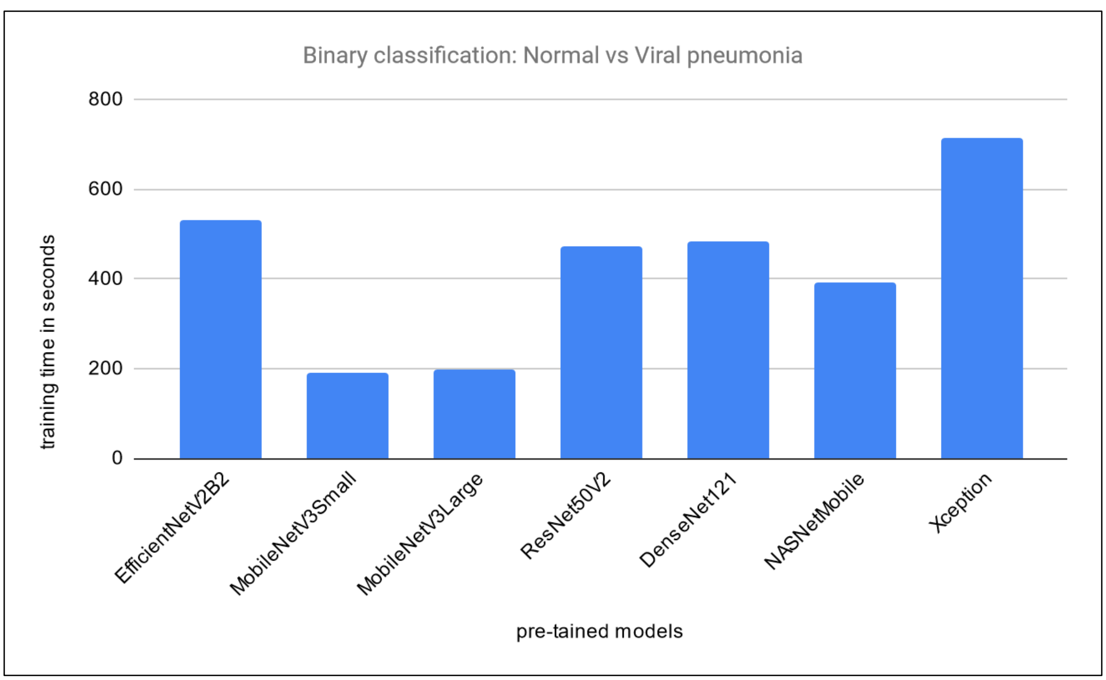


Figure 4. 5. Normal vs viral pneumonia model training time

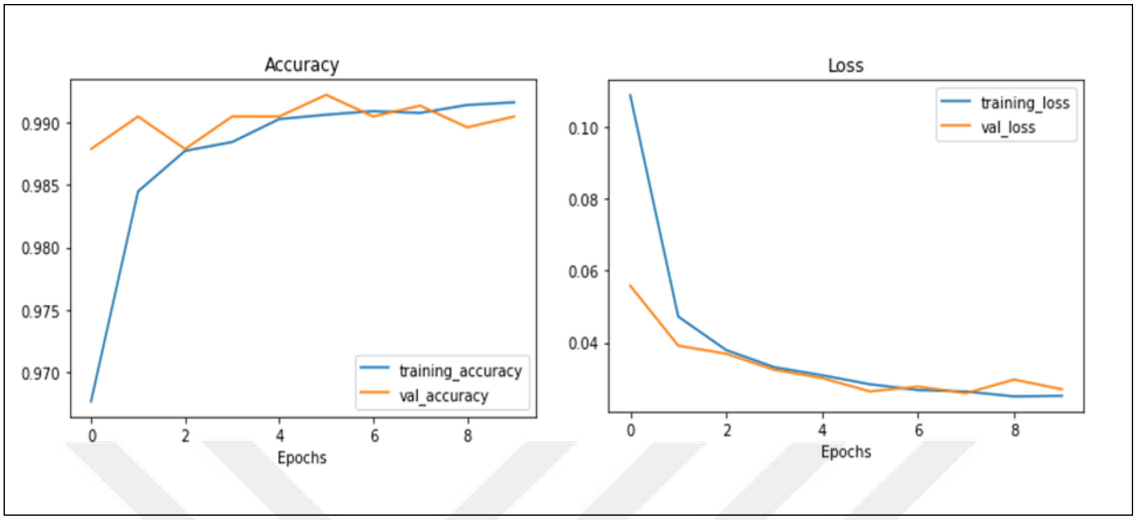


Figure 4. 6. Training accuracy and loss curves of the EfficientNetV2B2

4.1.3. Multi-class classification: normal vs viral pneumonia vs COVID-19

Table 4.3 exhibits the results generated by the pre-trained CNNs in the classification of normal, COVID-19, and viral pneumonia. The accuracy value of each of the neural nets was ≥ 90%. The precision, recall and F1-score of each model was ≥ 89%. The top performing CNN was the MobileNetV3Large which produced an accuracy, precision, recall and F1-score values ≥ 95%. The second best performing model was EfficeintNetV2B2 which registered accuracy, precision, recall and F1-Score values ≥ 94%. Figure 4.7. illustrates the confusion matrix of the MobileNetV3Large. Out of 269 viral pneumonia chest X-ray images, 260 were correctly classified as viral pneumonia positive and 9 were erroneously categorized as normal. Out of 2039 normal chest X-ray images, 1981 were correctly categorized as normal, 9 were incorrectly identified as viral pneumonia positive, and 49 were erroneously classified as COVID-19 positive. Out of 724 chest X-ray images, 665 were accurately categorized as COVID-19 positive, 58 were misclassified as normal and only 1 was inaccurately categorized as viral pneumonia positive. Figure 4.8. is a graph showing the pre-trained CNNs training time with MobileNetV3Small recording the shortest time of 275 seconds, followed closely by MobileNetV3Large with a training time of 283 seconds. Xception had the longest training time of 1055 seconds. Figure 4.9. displays the training accuracy and loss curves of the MobileNetV3Large in this classification category.

Table 4. 3. Performance metrics values of MobileNetV3Large

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	95	95	94	94
MobileNetV3Small	94	93	93	93
MobileNetV3Large	96	95	95	95
ResNet50V2	93	91	93	92
DenseNet121	90	89	89	89
NASNetMobile	91	89	90	89
Xception	92	91	90	91

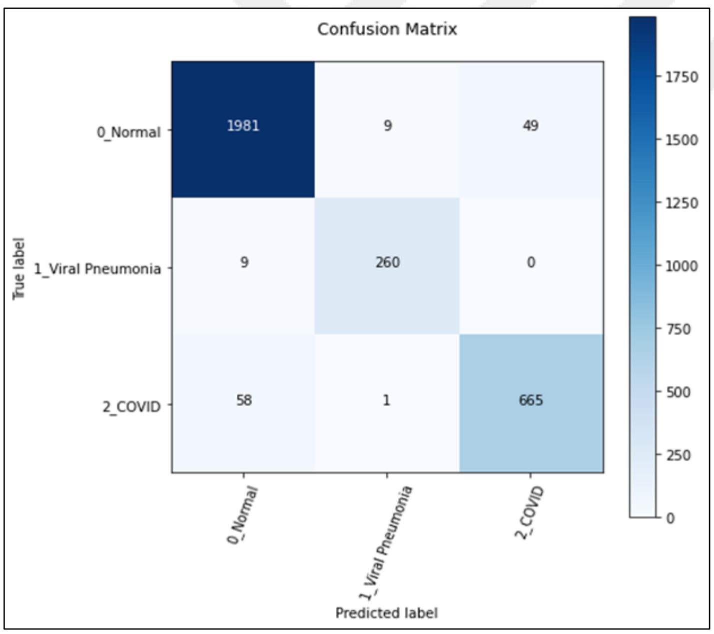


Figure 4. 7. Confusion matrix of MobileNetV3Large

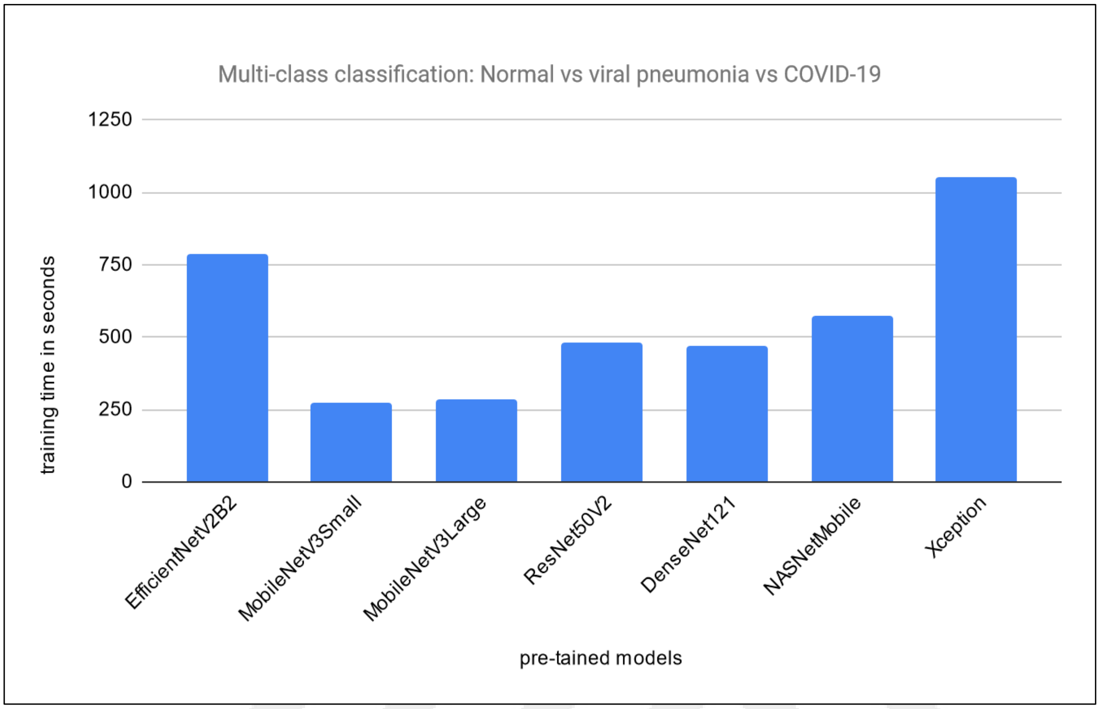


Figure 4. 8. Multi-class classification: normal vs viral pneumonia vs COVID-19 model training time.

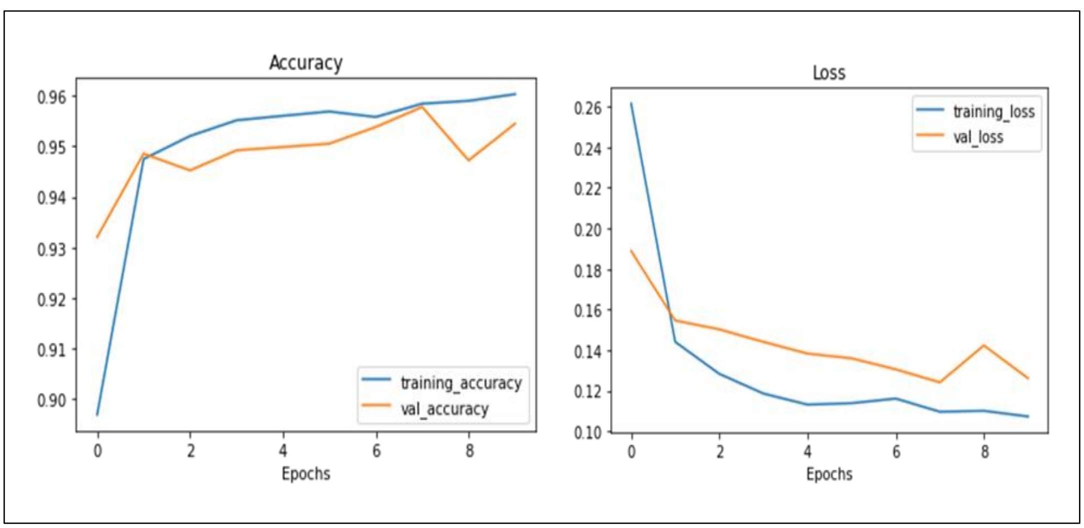


Figure 4. 9. Accuracy and loss curves of MobileNetV3Large

4.1.4. Multi-class classification: normal vs viral pneumonia vs COVID-19 vs lung opacity

Table 4.4. exhibits the models’ performances for the four class classification. The accuracy, recall, precision and F1-Score value of each of the pre-trained models was ≥ 81%. In this classification category, the EffcientNetV2B2 and MobileNetV3Large produced nearly similar performance outcomes. The two neural nets had an identical accuracy score of 89%. However, EfficientNetV2B2 had a slightly better F1-Score by 1%. Figure 4.10. shows the confusion matrix of the EfficientNetV2B2. Out of 269 viral pneumonia chest X-ray images, 254 were correctly classified as viral pneumonia positive, 14 misclassified as normal, and 1 misclassified as COVID-19 positive. Out of 2039 normal chest X-ray images, 1881 were correctly identified as normal, 12 misclassified as viral pneumonia positive, 20 misclassified as COVID-19 positive and 126 were misclassified as lung opacity. Out of 724 COVID-19 X-ray images, 635 were accurately categorized as COVID-19 positive, 62 misclassified as normal, 1 misclassified as viral pneumonia positive and 26 misclassified as lung opacity. Out of 1203 lung opacity chest X-ray images, 998 were correctly classified as lung opacity, 186 misclassified as normal, and 18 misclassified as COVID-19 positive and 1 was wrongly categorized as viral pneumonia positive. Figure 4.11. illustrates the time taken to train the model in this classification task. Xception had the worst training time with 1380 seconds followed by EfficientNetV2B2 which had a training time of 1033 seconds. MobileNetV3Small and MobileNetV3Large had the best training time with a difference of 10 seconds between the two models. MobileNetV3Small registered a training time of 362 seconds and MobileNetV3Large recorded a time of 372 seconds. Figure 4.12. shows the training accuracy and loss curves of the EfficientNetV2B2 in the normal vs viral pneumonia vs COVID-19 vs lung opacity classification category.

Table 4. 4. The accuracy, precision, recall, and F1-score values of the DenseNet121

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	89	91	89	90
MobileNetV3Small	88	89	88	88
MobileNetV3Large	89	89	90	89
ResNet50V2	87	88	87	87
DenseNet121	85	86	84	85
NASNetMobile	81	82	81	81
Xception	86	88	86	87

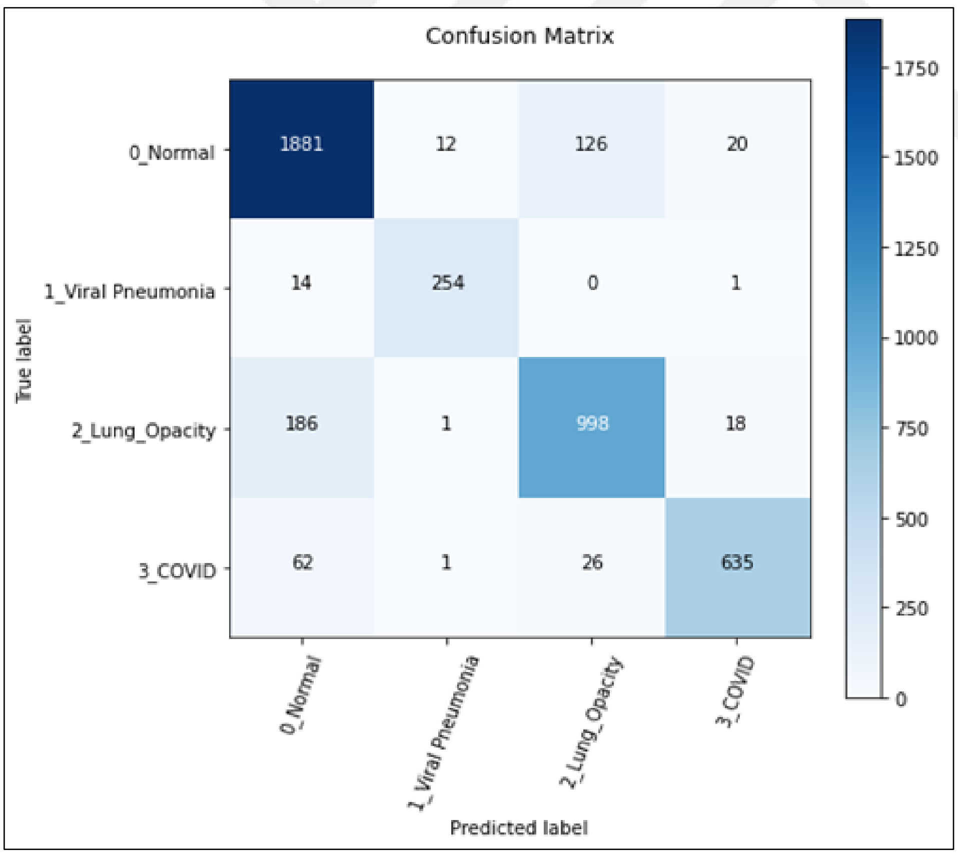


Figure 4. 10. Confusion matrix of MobileNetV3Large

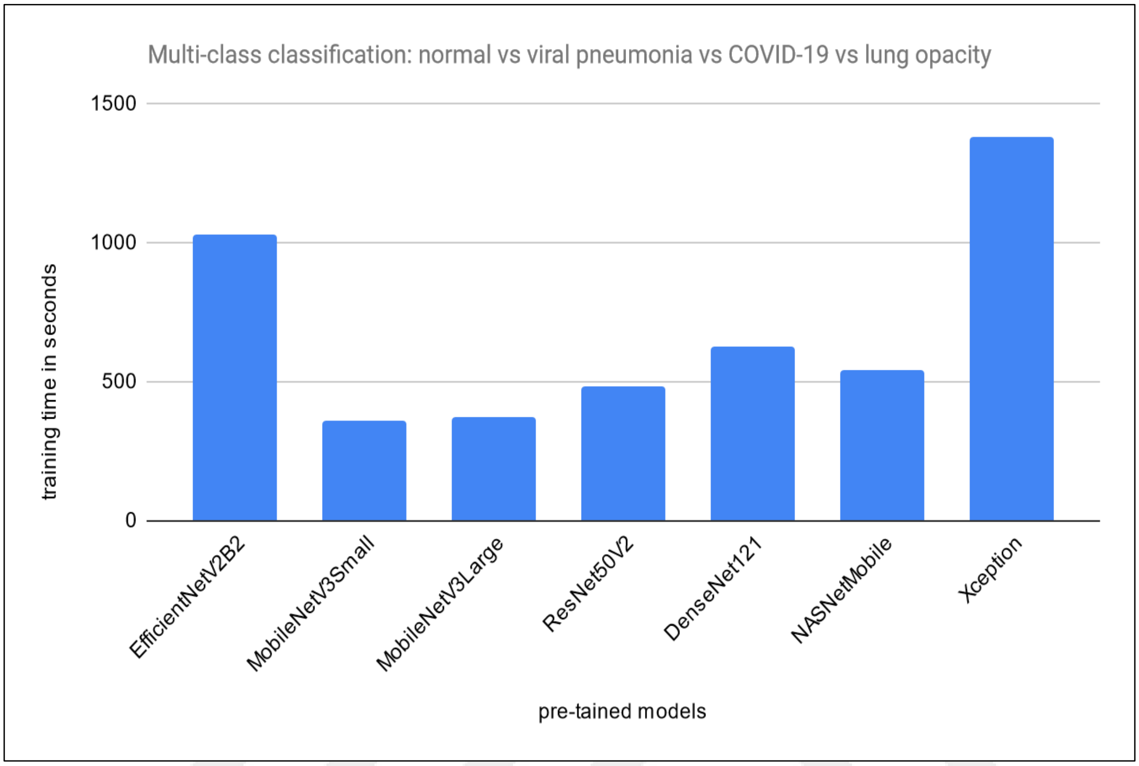


Figure 4. 11. Multi-class classification: normal vs viral pneumonia vs COVID-19 vs lung opacity model training time

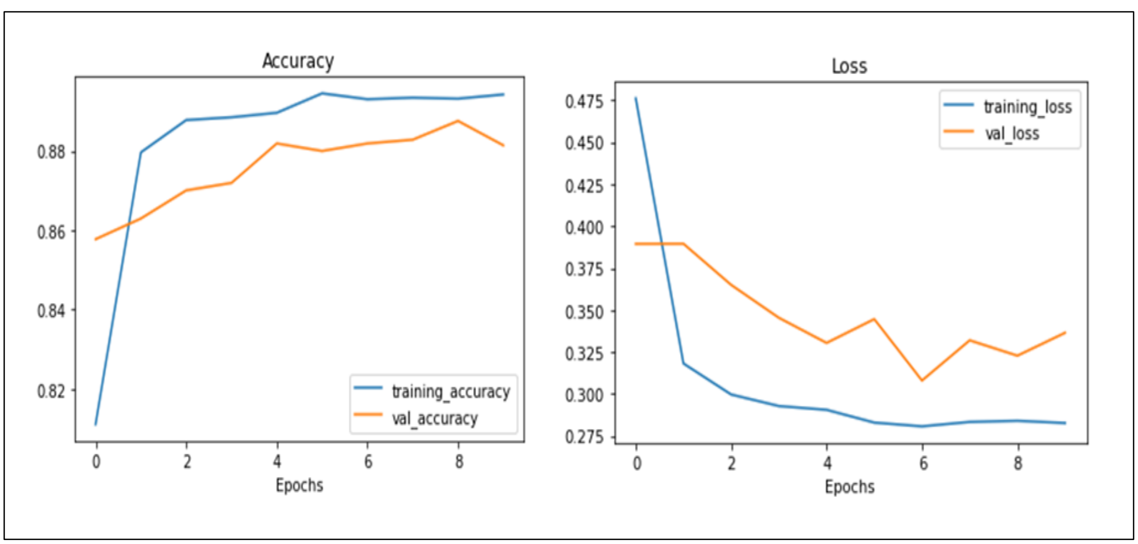


Figure 4. 12. Accuracy and loss curves of MobileNetV3Large

4.2. Pre-trained models as feature extractors and support vector machine (SVM) as classifier

In this model training and classification category, all the layers of the pre-trained model with the exception of the top most classification layer are frozen. The top most layer of the pre-trained model is replaced with a support vector machine which performs the classification task, whereas the frozen layers act as feature extractors.

4.2.1. Binary classification: normal vs COVID-19

Table 4.5. shows the performance outcomes of the pre-trained neural nets in the classification of normal and COVID-19 chest X-rays. Each of the pre-trained models registered an accuracy value ≥ 93%, precision value ≥ 94%, recall value ≥ 87% and F1-score ≥ 90%. The EfficientNetV2B2 neural net was the best performing model with an accuracy, precision, recall and F1-Score of 97%, 97%, 94%, 95% respectively. Figure 4.13 shows the confusion matrix of the EfficientNetV2B2 model. Out of the 724 COVID-19 chest X-ray images 635 were correctly categorized as COVID-19 positive and 89 were incorrectly categorized as normal. Out 2039 normal chest X-ray images 2032 were correctly identified as normal and only 7 were wrongly classified as COVID-19 positive. Figure 4.14. shows the time taken to train the pre-trained CNNs in the normal vs COVID-19 classification. Figure 4.15. presents the training accuracy and loss curves of the EfficientNetV2B2 in this classification category. MobileNetV3Small took the least amount of time to train with 187 seconds. MobileNetV3Large recorded the second shortest training time with 192 seconds. The Xception model took the longest time to train with 701 seconds.

Table 4. 5. Performance metrics values of EfficientNetV2B2

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	97	97	94	95
MobileNetV3Small	95	96	91	93
MobileNetV3Large	96	97	92	94
ResNet50V2	94	96	90	92
DenseNet121	94	95	89	91
NASNetMobile	93	94	87	90
Xception	93	95	88	91

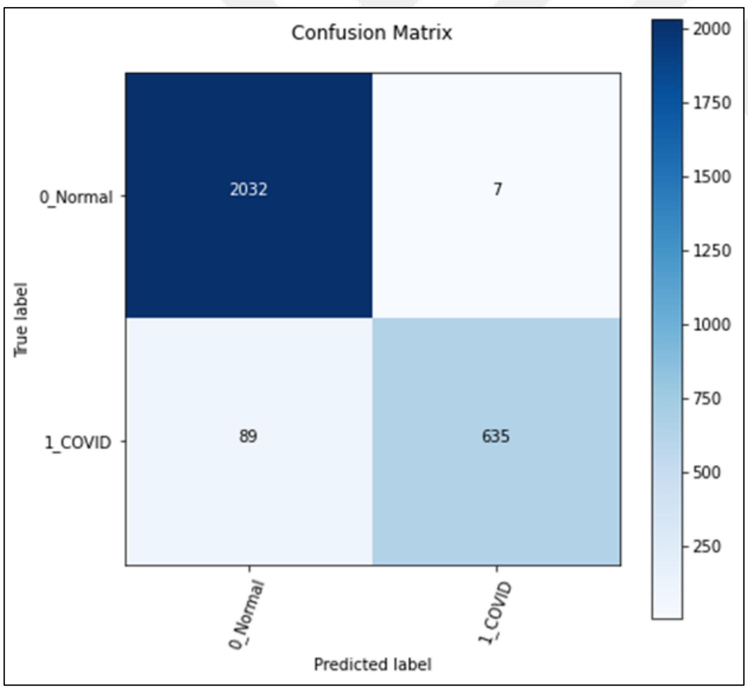


Figure 4. 13. Confusion matrix of EfficientNetV2B2

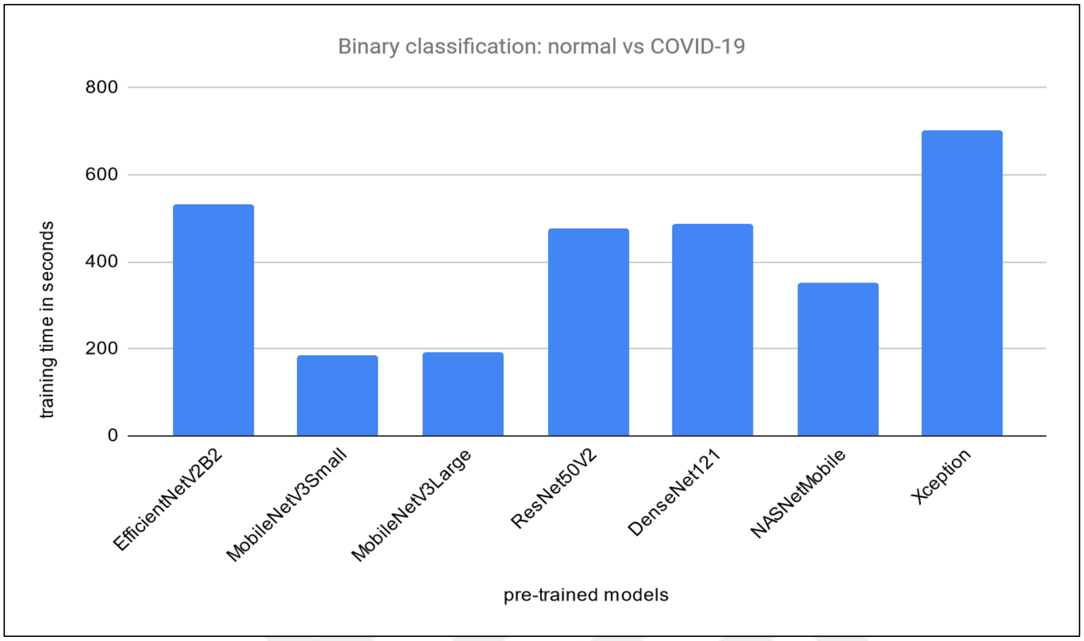


Figure 4. 14. Normal vs COVID-19 model training time

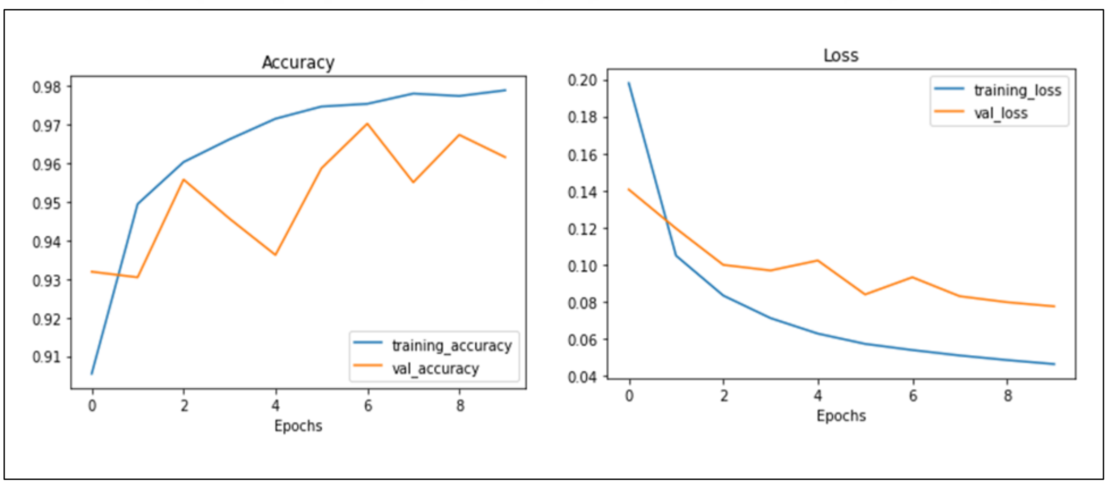


Figure 4. 15. Accuracy and loss curves of EfficientNetV2B2

4.2.2. Binary classification: normal vs viral pneumonia

Table 4.6. illustrates results generated by the pre-trained CNNs in the identification of normal and viral pneumonia chest X-rays. All the CNNs generated accuracy and precision scores ≥ 98%, recall scores ≥ 90% and F1-Scores ≥ 94%. The EffcientNetV2B2, MobileNetV3Small, MobileNetV3Large and ResNet50V2 all recorded almost similar metrics scores. All the four

models had the same accuracy score of 99% with the only difference being in the precision, recall and F1-Score values which differed with a maximum of 2%. The NASNetMobile and Xception CNNs were the models that had the least performance metric values when compared to the other models. Figure 4.16. shows the confusion matrix of the EffcientNetV2B2. Out of 269 viral pneumonia chest X-ray images, 243 were correctly categorized as viral pneumonia positive and 26 were misclassified as normal X-ray images. For the normal chest X-ray images, only a single chest X-ray image was erroneously categorized as viral pneumonia positive. Figure 4.17. displays the time taken to train the pre-trained neural nets in the classification of normal and viral pneumonia chest X-rays. MobileNetV3Small had the least training time of 186 second. Xception had the longest training time of 700 seconds. Figure 4.18. displays the accuracy and loss curves when training the EfficientNetV2B2 neural net in this classification category.

Table 4. 6. Performance metrics values of the pre-trained CNNs

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
EfficientNetV2B2	99	99	95	97
MobileNetV3Small	99	98	97	98
MobileNetV3Large	99	99	97	98
ResNet50V2	99	98	98	98
DenseNet121	99	99	95	97
NASNetMobile	98	98	90	94
Xception	98	99	90	94

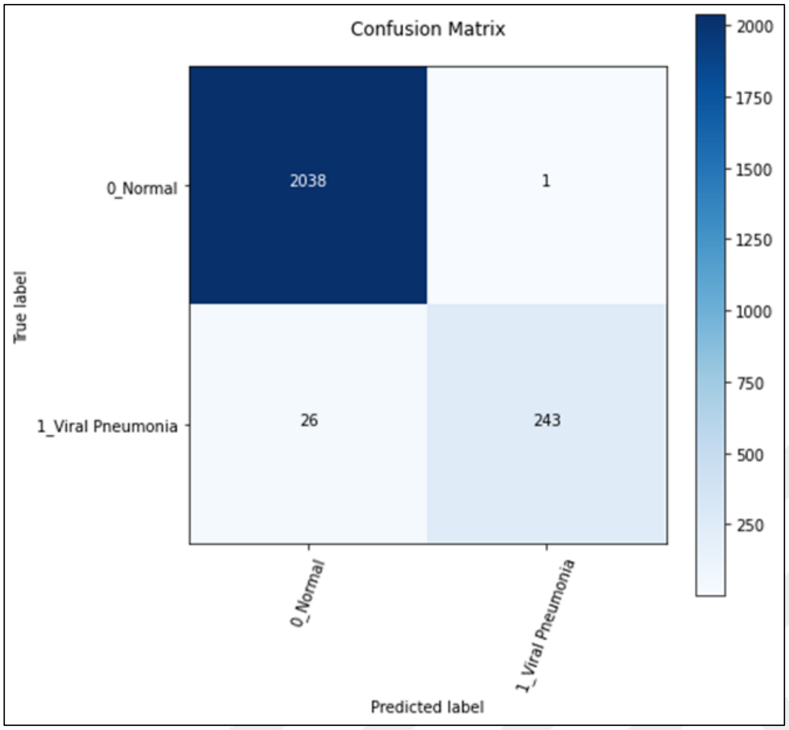


Figure 4. 16. Confusion matrix of EfficientNetV2B2

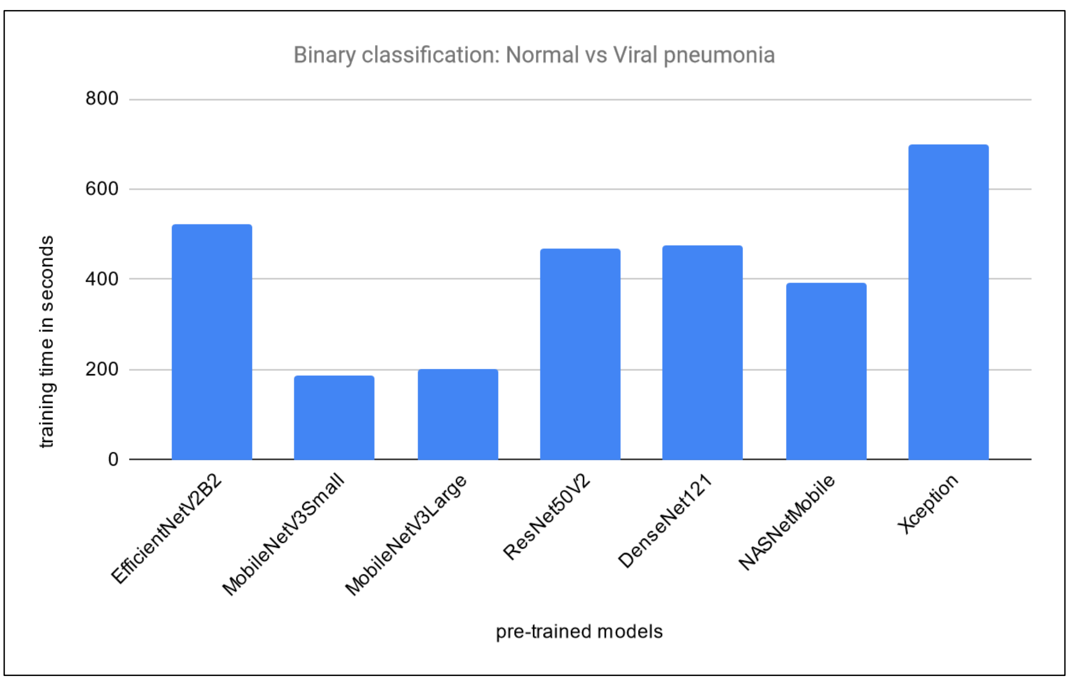


Figure 4. 17. Normal vs viral pneumonia model training time

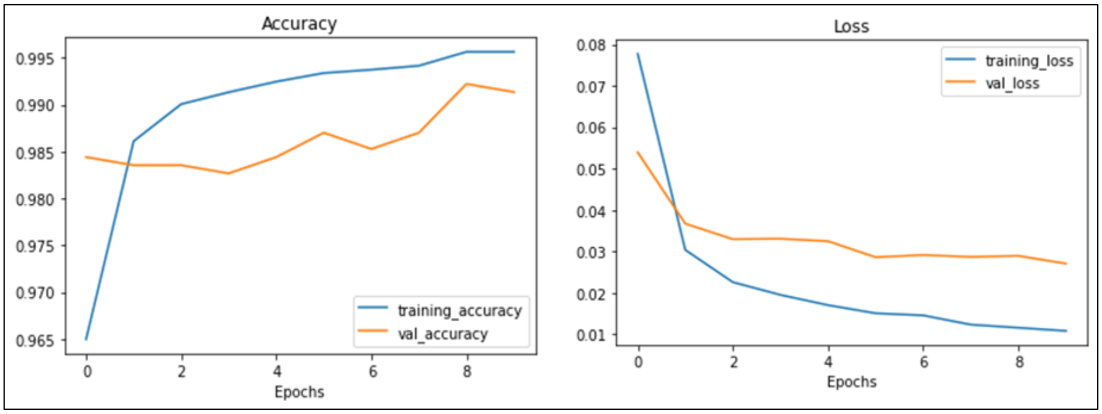


Figure 4. 18. Accuracy and loss curves of EfficientNetV2B2

4.2.3. Multi-class classification: normal vs viral pneumonia vs COVID-19

Table 4.7. shows the performance outcomes generated by the pre-trained models in the multi-class classification of normal, COVID-19, and viral pneumonia. The accuracy, precision, recall and F1-Score values given by the CNNs were ≥ 91%. MobileNetV3Large, MobileNetV3Small and EfficientNetV2B2 were the best performing CNNs. However, the MobileNetV3Large model recorded a slightly better accuracy and F1-Score value i.e 1% higher than the other two models. This made it the top performing CNN in this classification category. Figure 4.19. illustrates the confusion matrix of the MobileNetV3Large. Out of 269 viral pneumonia chest X-ray images, 265 were rightly classified as viral pneumonia positive and 4 were erroneously categorized as normal. Out of 2039 normal chest X-ray images, 1968 were rightly categorized as normal, 9 were incorrectly identified as viral pneumonia positive, and 62 were incorrectly classified as COVID-19 positive. Out of 724 chest X-ray images, 693 were accurately categorized as COVID-19 positive, 30 were misclassified as normal and only 1 was inaccurately categorized as viral pneumonia positive. Figure 4.20. shows the time taken to train the pre-trained models The MobileNetV3Small took the shortest time to train with 277 seconds. Xception model took the longest time to train with 1043 seconds. Figure 4.21. displays the training accuracy and loss curves of the MobileNetV3Large in this classification category

Table 4. 7. Performance metrics values of the pre-trained CNNs

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	96	96	95	95
MobileNetV3Small	95	94	94	94
MobileNetV3Large	97	95	97	96
ResNet50V2	94	93	92	92
DenseNet121	93	93	92	92
NASNetMobile	93	91	92	91
Xception	92	91	91	91

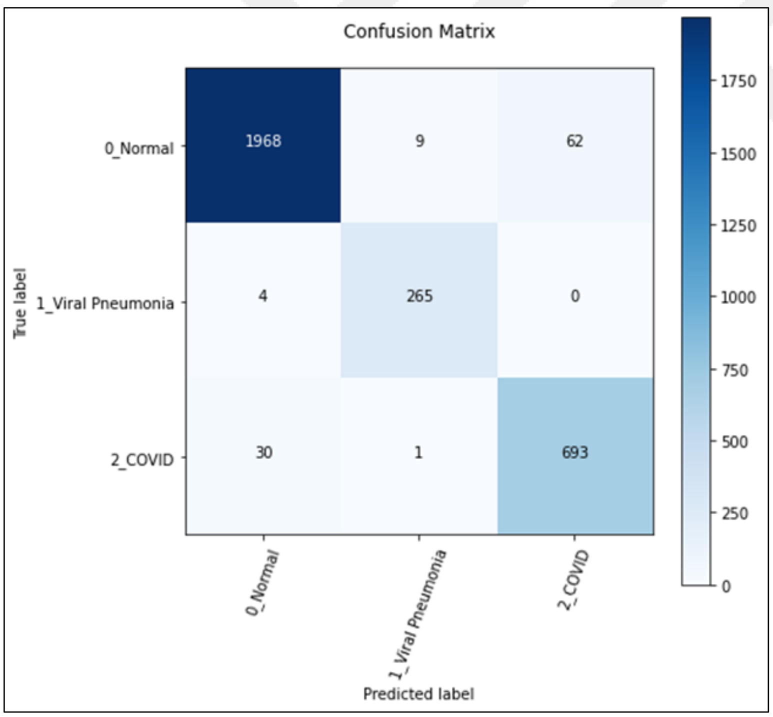


Figure 4. 19. Confusion Matrix of MobileNetV3Large

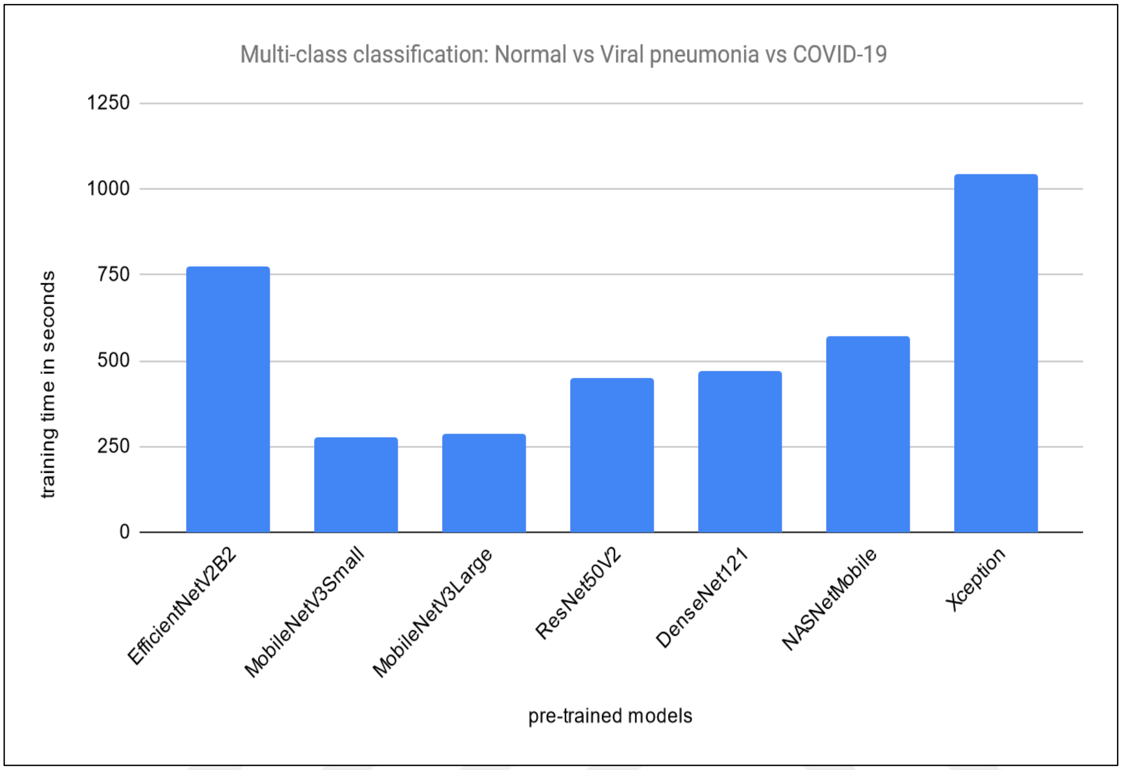


Figure 4. 20. Multi-class classification: normal vs viral pneumonia vs COVID-19 model training time

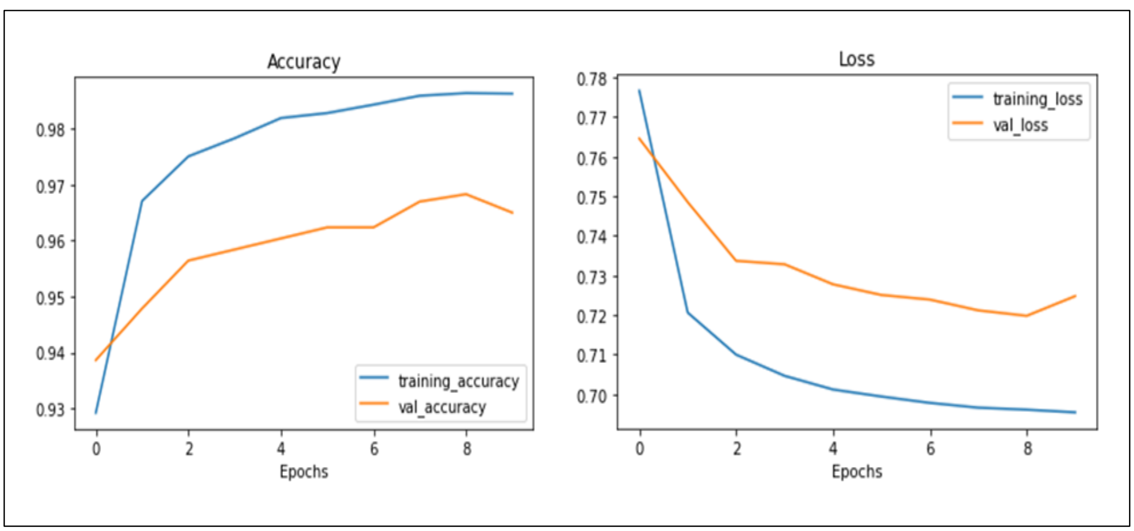


Figure 4. 21. Accuracy and loss curves of MobileNetV3Large

4.2.4. Multi-class classification: normal vs viral pneumonia vs COVID-19 vs lung opacity

Table 4.8. exhibits the models' performance evaluation outcomes for the seven pre-trained models. The accuracy, recall, precision and F1-Score values of each of the pre-trained models was ≥ 81%. The EffcientNetV2B2 and MobileNetV3Large generated closely matched performance outcomes. There was only a 1% difference in the recall and precision metrics. EffcientNetV2B2 had the highest precision value whereas MobileNetV3Large had the highest recall value. Figure 4.22. shows the confusion matrix of the MobileNetV3Large neural net. Out of 269 viral pneumonia chest X-ray images, 261 were accurately classified as viral pneumonia positive, 6 misclassified as normal, and 2 misclassified as COVID-19 positive. Out of 2039 normal chest X-ray images, 1834 were correctly identified as normal, 14 incorrectly classified as viral pneumonia positive, 15 misclassified as COVID-19 positive and 176 were misclassified as lung opacity. Out of 724 COVID-19 chest X-ray images, 647 were accurately categorized as COVID-19 positive, 38 misclassified as normal, 1 misclassified as viral pneumonia positive and 38 wrongly classified as lung opacity. Out of 1203 lung opacity chest X-ray images, 1086 were rightly classified as lung opacity, 95 misclassified as normal, and 21 misclassified as COVID-19 positive and 1 was incorrectly categorized as viral pneumonia positive. Figure 4.23. illustrates the time taken to train the models in this classification task. Xception and EfficientNetV2B2 had the worst training time of 1035 seconds and 1394 seconds respectively. MobileNetV3Small had the least training time of 362 seconds. Figure 4.24. shows the training accuracy and loss curves of the MobileNetV3Large in the normal vs viral pneumonia vs COVID-19 vs lung opacity classification category.

Table 4. 8. Performance metrics values of the pre-trained CNNs

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	90	92	90	91
MobileNetV3Small	89	91	90	90
MobileNetV3Large	90	91	92	91
ResNet50V2	86	88	84	85
DenseNet121	87	90	87	88
NASNetMobile	83	84	81	82
Xception	86	87	87	87

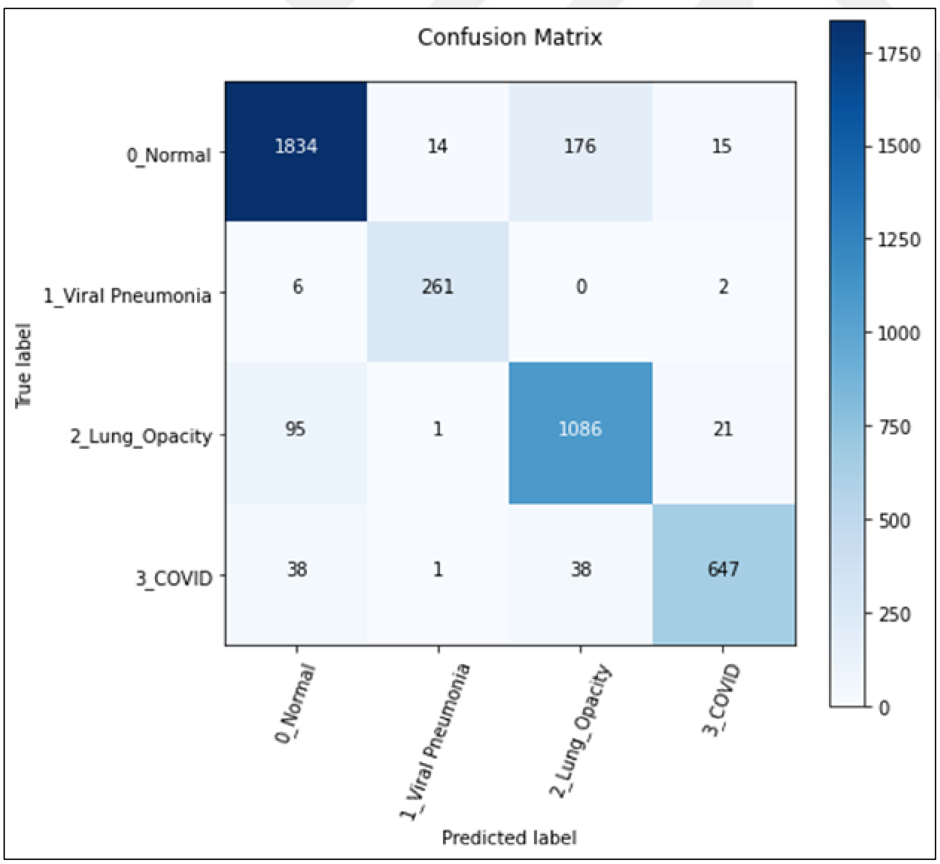


Figure 4. 22. Confusion Matrix of MobileNetV3Large

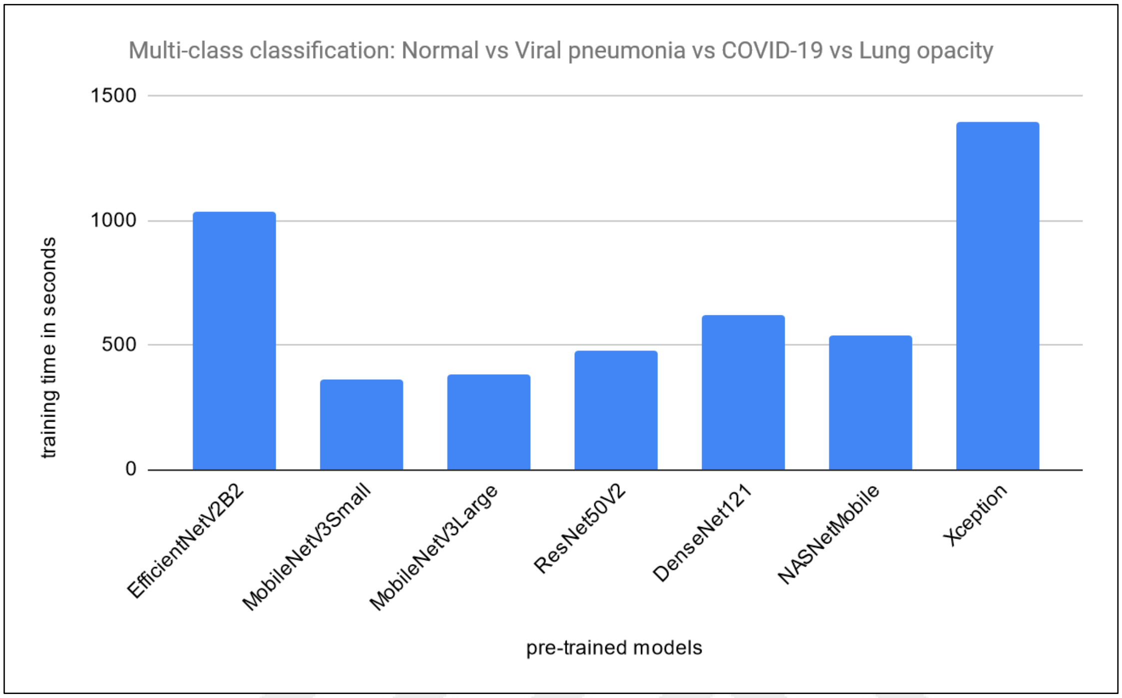


Figure 4. 23. Multi-class classification: normal vs viral pneumonia vs COVID-19 vs lung opacity model training time

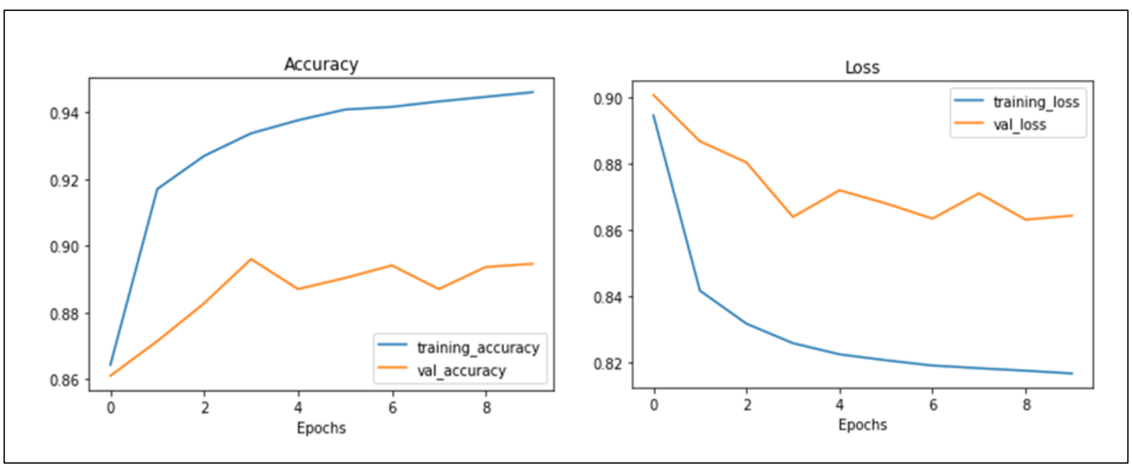


Figure 4. 24. The accuracy and loss curves of MobileNetV3Large

4.3. Model fine-tuning

In this model training and classification category, the deeper layers of the pre-trained model are frozen where as some of the shallow model layers are unfrozen during model training. The weights of the unfrozen layers are adjusted when the model is being trained, enabling the model to better learn features specific to the training dataset.

4.3.1. Binary classification: normal vs COVID-19

Table 4.9. presents the performance metrics of all the seven models in the identification normal and COVID-19 chest X-rays. All the models performed excellently well with most of the models producing accuracies ≥ 98%, precisions ≥ 97%, recalls ≥ 97% and F1-scores ≥ 97%. NASNetMobile was the worst performing model with the least accuracy score of 92% and recall value as low as 88%. The best performing model out of the seven was the EfficientNetV2B2 which had an accuracy, precision and F1-score of 99%. Figure 4.25. shows the confusion matrix of the model. Out of the 724 COVID-19 X-ray images 708 were correctly classified as COVID-19 positive and only 16 were misclassified as normal. Out 2039 normal X-ray images 2025 were correctly identified as normal and only 14 were misclassified as COVID-19 positive. Figure 4.26. is a graph displaying the time taken to train the pre-trained CNNs in the normal vs COVID-19 classification. The Xception neural net takes the highest amount of time to train with 840 seconds. MobileNetV3Small and MobileNetV3Large record the least training time with 211 seconds and 219 seconds respectively. Figure 4.27. shows the training accuracy and loss curves of the EfficientNetV2B2 in this classification category.

Table 4. 9. Performance metrics values of the pre-trained CNNs

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	99	99	99	99
MobileNetV3Small	98	97	97	97
MobileNetV3Large	99	99	98	98
ResNet50V2	98	98	97	98
DenseNet121	98	98	96	97
NASNetMobile	92	91	88	90
Xception	98	99	97	98

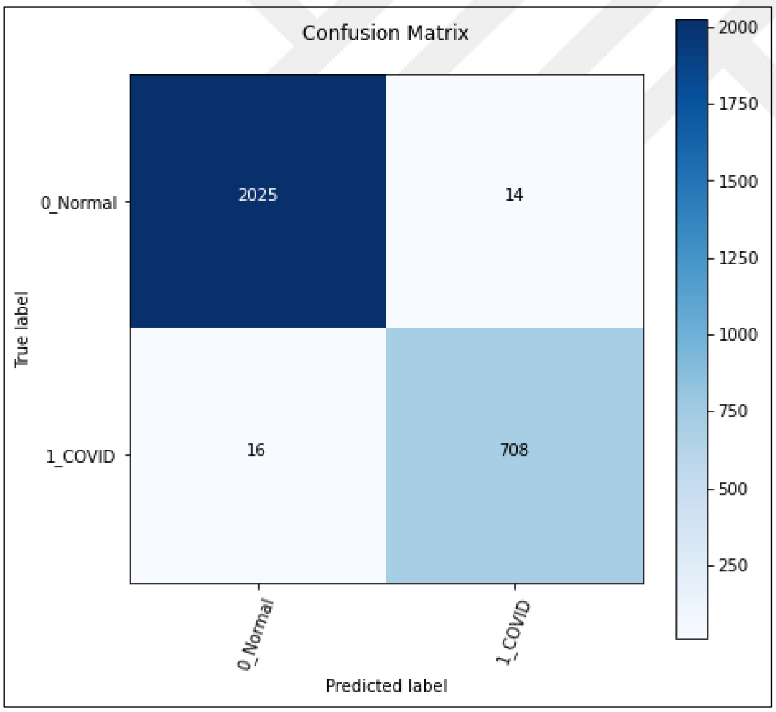


Figure 4. 25. Confusion matrix of EfficientNetV2B2

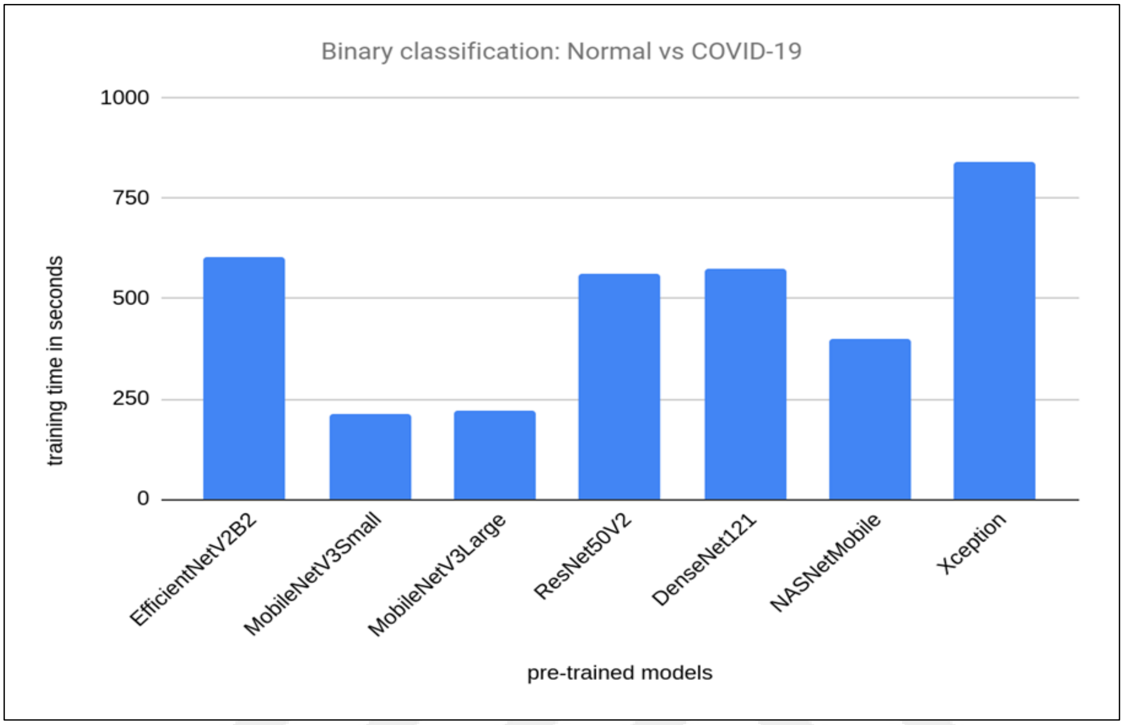


Figure 4. 26. Normal vs COVID-19 model training time

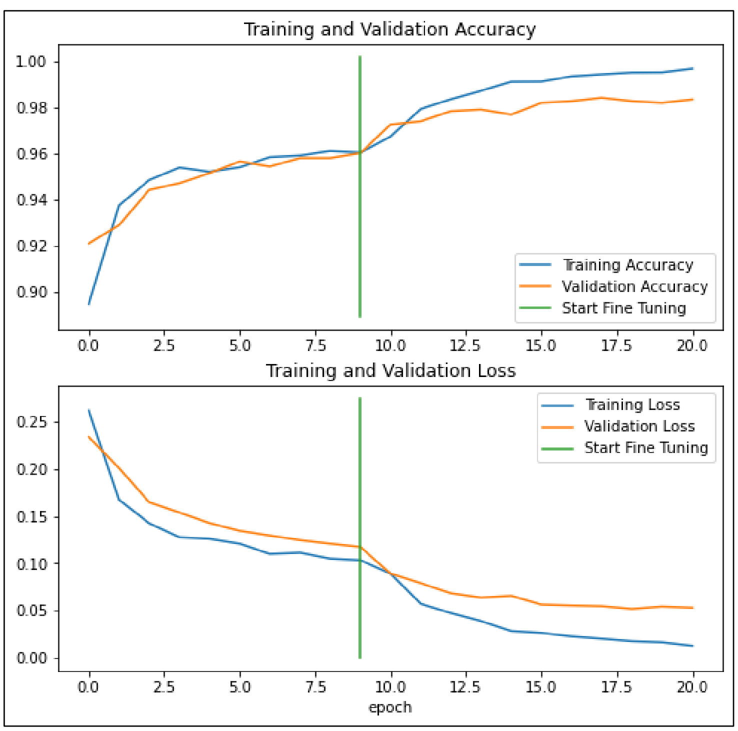


Figure 4. 27. Accuracy and loss curves of EfficientNetV2B2

4.3.2. Binary classification: normal vs viral pneumonia

Table 4.10. illustrates how all the seven models performed in the identification normal and viral chest X-rays. All the models had an accuracy score of 99% except for Xception and NASNetMobile which had an accuracies of 98%. The precision and recall values for each of the models was ≥ 94%. The F1-Score value of each of the models was ≥ 94%. The EffcientNetV2B2 topped in three of the performance metrics i.e. accuracy, precision and F1-score with a value of 99%. Figure 4.28. shows the confusion matrix of the CNN. Out of 269 Viral pneumonia images, 259 were correctly identified as viral pneumonia positive, and only 10 were wrongly categorized as Normal. Figure 4.29. is a graph showing the training time of the neural nets in the classification of normal and viral pneumonia chest X-rays. MobileNetV3Small recorded the shortest training time of 193 seconds followed by MobileNetV3Large which registered a training time of 208 seconds respectively. Xception took the longest time to train with 835 seconds. Figure 4.30. presents the training accuracy and loss curves of the EfficientNetV2B2 in this classification category.

Table 4. 10. Performance metrics values for the pre-trained CNNs

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	99	99	98	99
MobileNetV3Small	99	97	99	98
MobileNetV3Large	99	98	98	98
ResNet50V2	99	98	97	98
DenseNet121	99	97	98	98
NASNetMobile	98	94	94	94
Xception	98	99	97	98

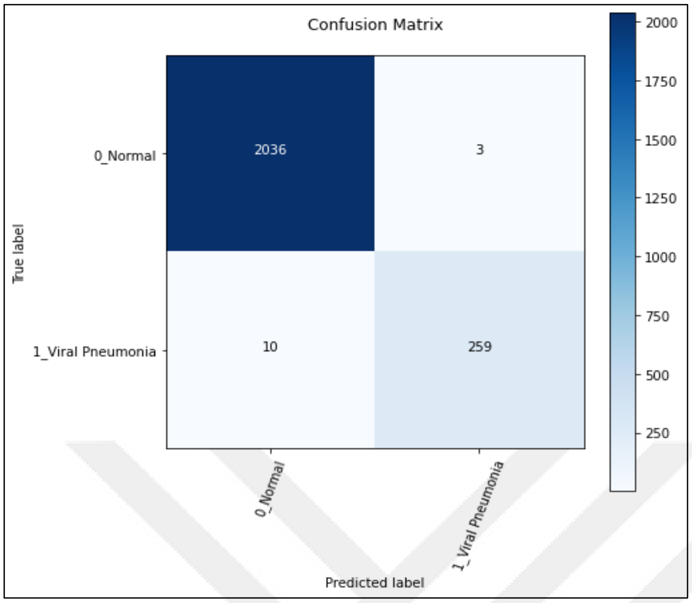


Figure 4. 28. Confusion matrix of EfficientNetV2B2

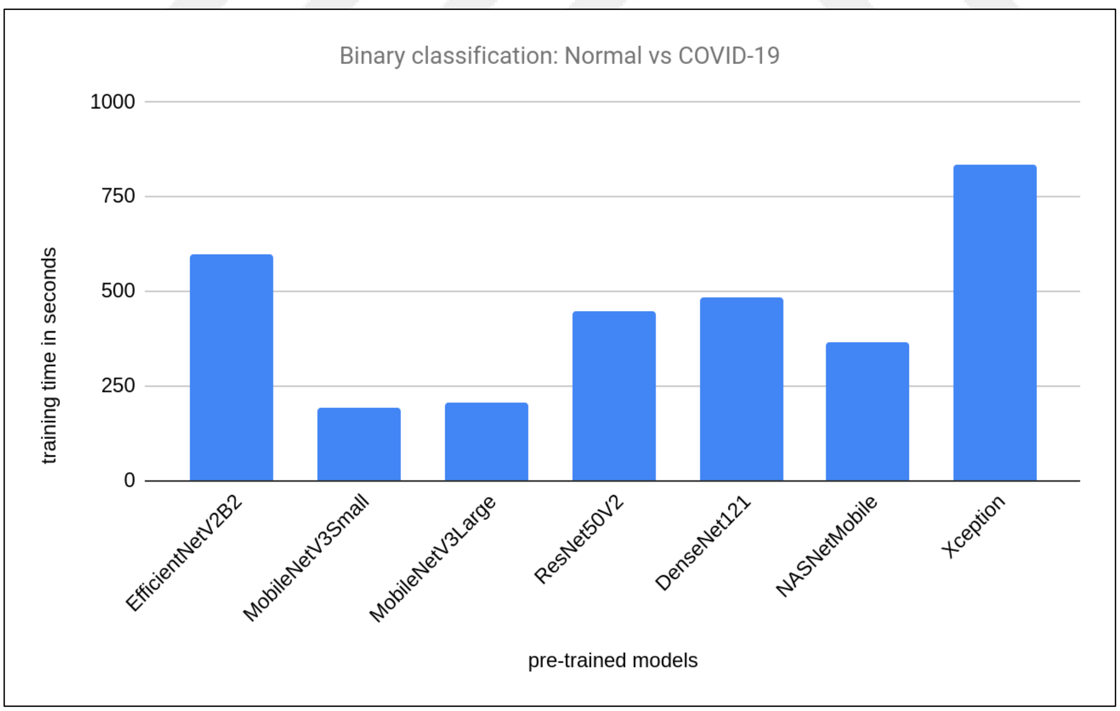


Figure 4. 29. Normal vs viral pneumonia model training time

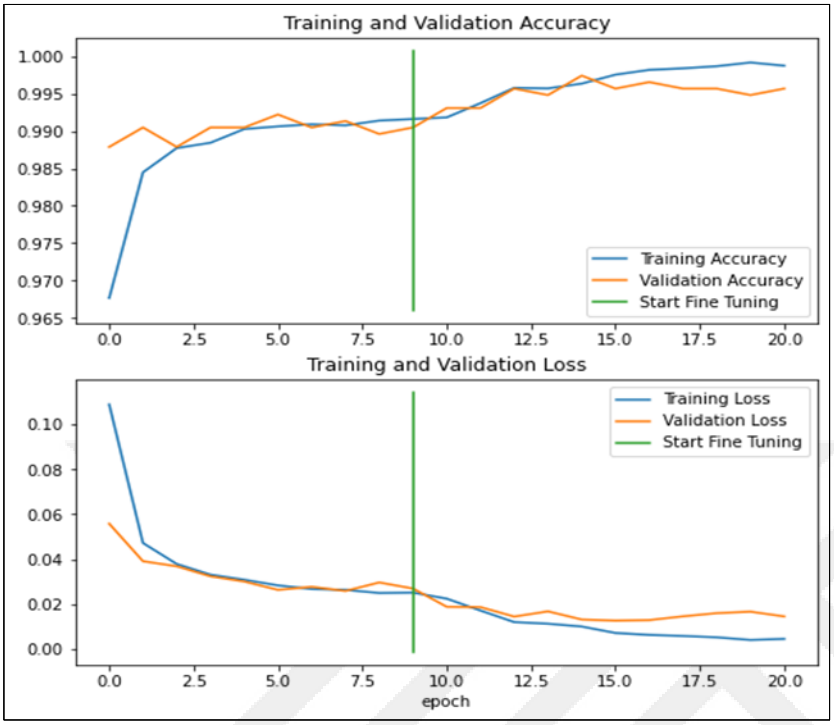


Figure 4. 30. Accuracy and loss curves of EfficientNetV2B2

4.3.3. Multi-class classification: normal vs viral pneumonia vs COVID-19

Table 4.11. exhibits the results generated by the pre-trained CNNs in the classification of normal, COVID-19, and viral pneumonia. The accuracy of six of the seven CNNs was ≥ 97%. The precision and F1-score of six of the seven models was ≥ 96%. The recall value of six of the seven models was ≥ 95%. The top performing CNN was the MobileNetV3Large which produced an accuracy, recall and F1-score of 99% and a precision of 98%. Figure 4.31. illustrates the confusion matrix of the MobileNetV3Large. Out of 269 viral pneumonia X-ray images, 267 were correctly classified as viral pneumonia positive and only 2 were incorrectly categorized as normal. Out of 2039 normal X-ray images, 2021 were correctly classified as normal, 7 were incorrectly identified as Viral pneumonia positive, and 11 were erroneously classified as COVID-19 positive. Out of 724 X-ray images, 709 were accurately categorized as COVID-19 positive, and only 15 were inaccurately categorized as normal. Figure 4.32. is a graph showing the pre-trained CNNs training time with MobileNetV3Small registering the best time of 303 seconds. Xception neural net had the worst training time of 1241 seconds. Figure

4.33. Displays the training accuracy and loss curves of the MobileNetV3Large in this classification category.

Table 4. 11. Performance metrics values for the pre-trained CNNs

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	98	97	97	97
MobileNetV3Small	97	96	97	97
MobileNetV3Large	99	98	99	99
ResNet50V2	98	97	97	97
DenseNet121	97	96	97	97
NASNetMobile	90	90	87	88
Xception	97	97	95	96

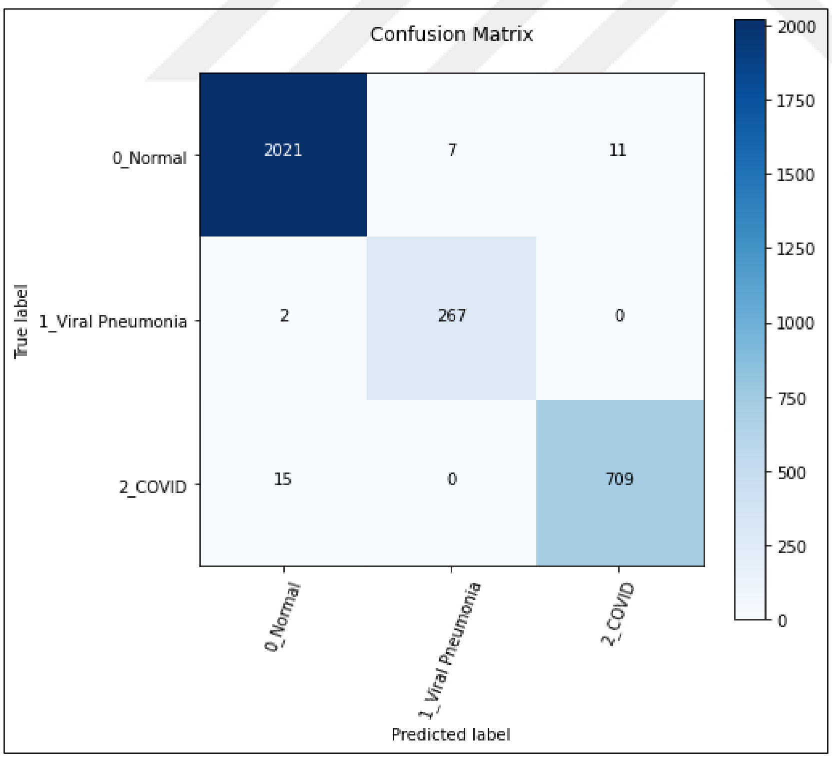


Figure 4. 31. Confusion matrix of MobileNetV3Large

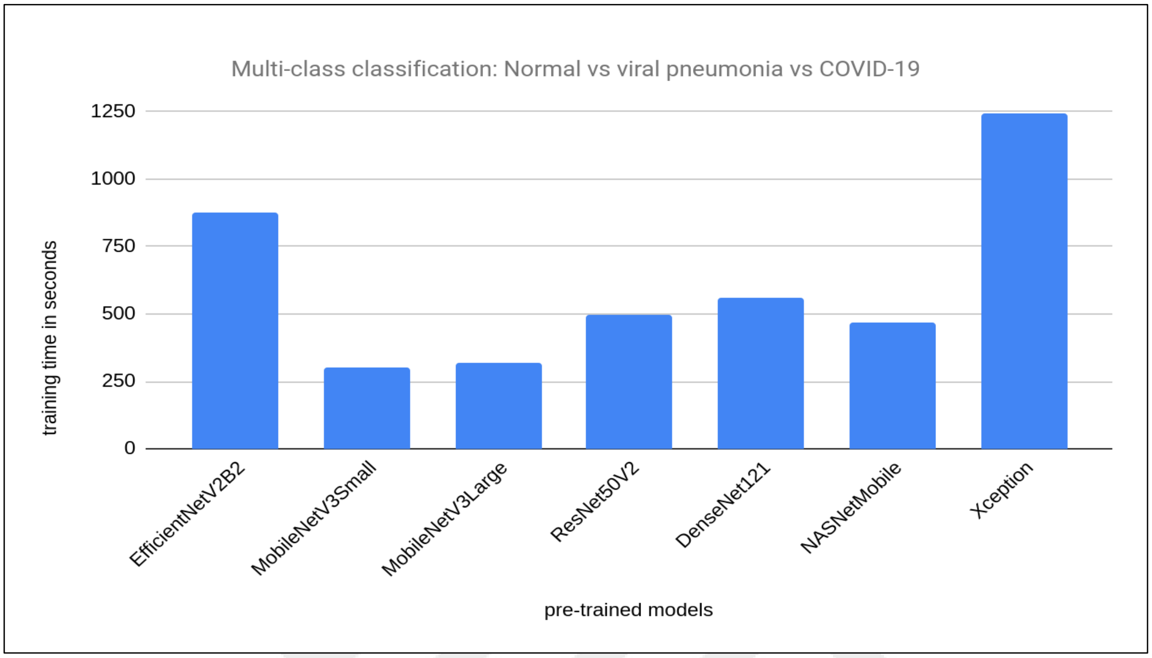


Figure 4. 32. Multi-class classification: normal vs viral pneumonia vs COVID-19 model training time.

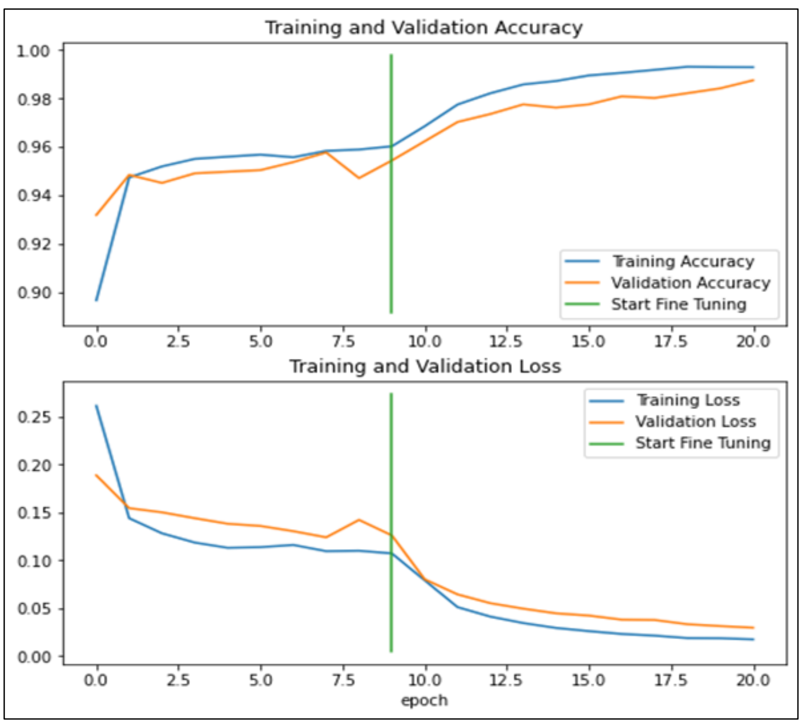


Figure 4. 33. Accuracy and loss curves of MobileNetV3Large

4.3.4. Multi-class classification: normal vs viral pneumonia vs COVID-19 vs lung opacity

Table 4.12. exhibits the results of the 4-class model classification. The accuracy value of six of the seven pre-trained models was ≥ 91%. The recall and F1-score of six of the seven models was ≥ 92% and the precision was ≥ 93%. The worst performing model was the NASNetMobile which had accuracy level low as 82% and recall score as low as 77%. In this classification category, the EffcientNetV2B2, MobileNetV3Large and DeseNet121 produced nearly the same results. The models had a similar precision, recall and F1-score of 95%. The only difference was in the accuracy where the DenseNet121 scored 94% which was 1% higher than the other two models. Figure 4.34. shows the confusion matrix of the DenseNet121. Out of 269 viral pneumonia X-ray images, 261 were correctly classified as viral pneumonia positive, 7 misclassified as normal, and 1 misclassified as lung opacity. Out of 2039 normal X-ray images, 1931 were correctly identified as normal, 6 misclassified as viral pneumonia positive, 18 misclassified as COVID-19 positive and 84 were misclassified as lung opacity. Out of 724 COVID-19 X-ray images, 709 were accurately categorized as COVID-19 positive, 13 misclassified as normal, 1 misclassified as viral pneumonia positive and 1 misclassified as lung opacity. Out of 1203 lung opacity X-ray images, 1075 were correctly classified as lung opacity, 110 misclassified as normal, and 18 misclassified as COVID-19 positive. Figure 4.35. illustrates the time taken to train the model in the classification of normal, COVID-19, viral pneumonia and lung opacity. Xception had the worst training time with 1644 seconds followed by EfficientNetV2B2 which had a training time of 1172 seconds. MobileNetV3Small and MobileNetV3Large had the best training time of 418 and 421 seconds respectively. Figure 4.36. shows the training accuracy and loss curves of the MobileNetV3Large in the normal vs viral pneumonia vs COVID-19 vs lung opacity classification category.

Table 4. 12. Performance metrics values for the pre-trained CNNs

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNetV2B2	93	95	95	95
MobileNetV3Small	91	93	92	92
MobileNetV3Large	93	95	95	95
ResNet50V2	92	93	93	93
DenseNet121	94	95	95	95
NASNetMobile	82	86	77	81
Xception	92	94	93	96

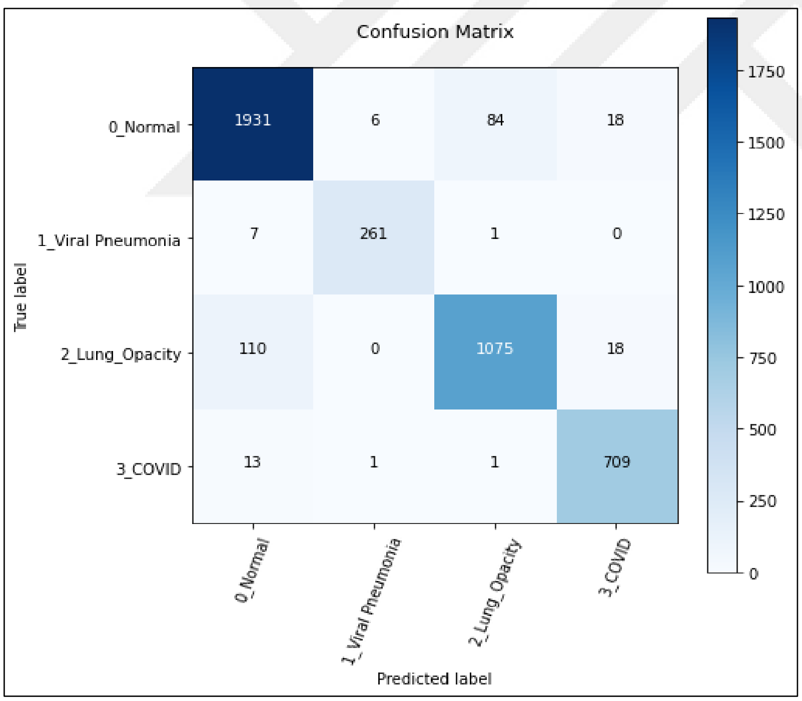


Figure 4. 34. Confusion matrix of MobileNetV3Large

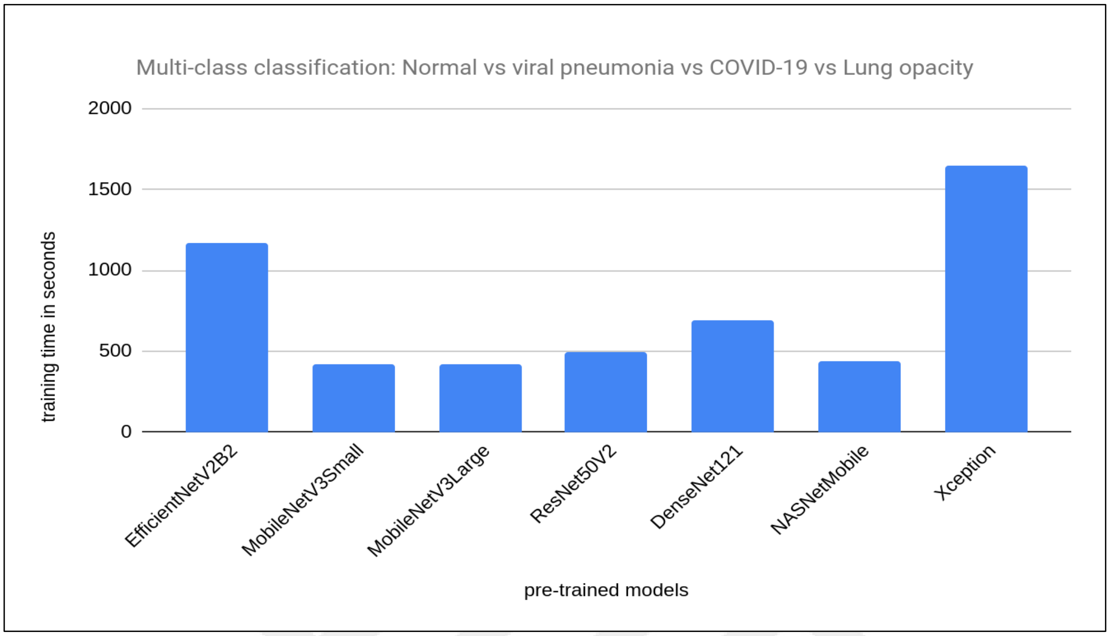


Figure 4. 35. Multi-class classification: normal vs viral pneumonia vs COVID-19 vs lung opacity model training time.

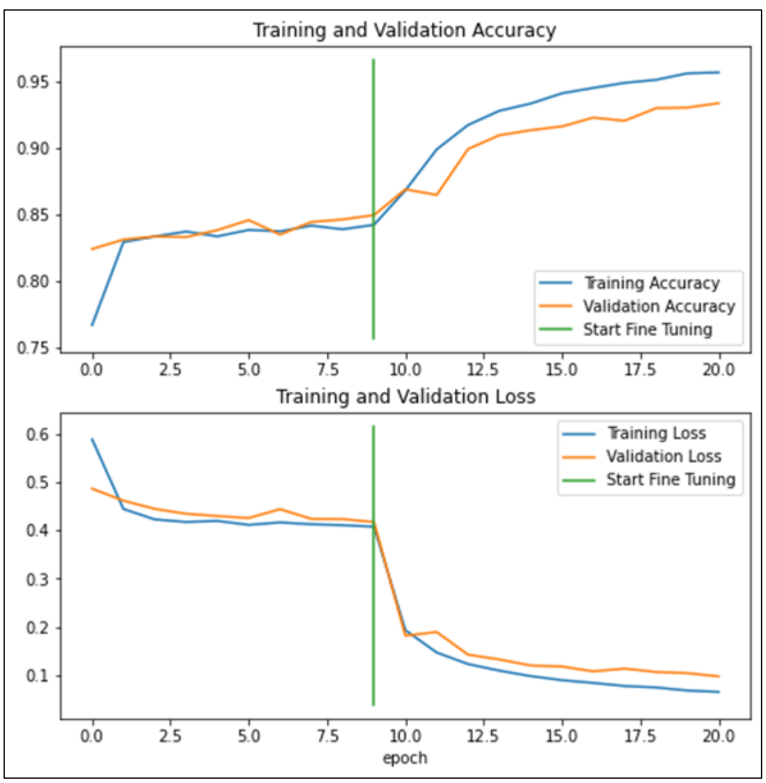


Figure 4. 36. Accuracy and loss curves of MobileNetV3Large

5. CONCLUSION

The results produced by the pre-trained neural nets in detection of viral pneumonia and COVID-19 patients were very promising. A majority of the models registered the least performance metric values when they were trained using feature extraction Transfer learning using a dense layer as the classification layer. There was a marginal increase in the metric values when the support vector machine was used to carry out the classification task. Fine-tuning transfer learning generated the best model performance metric values. However, the pre-trained models took the longest training time when trained using Fine-tuning transfer learning. For example, in the classification of normal and COVID-19 chest X-ray images. The MobileNetV3Small training time was 187 seconds for Feature extraction using a dense classification layer, 200 seconds for Feature extraction and an SVM classifier, and 211 seconds for Fine-tuning Transfer learning. The faster training time makes Feature extraction transfer learning suitable for conducting quick baseline experiments for determining models suitable for a particular task. After selecting the most appropriate model for a given task, the model can be fine-tuned to further enhance its performance. The longer training time in Fine-tuning can be attributed to the unfreezing of some of the model layers which increases the number of parameters to be updated during training.

The pre-trained models had better performance metrics in binary classification than in multi-class classification that consisted of more than two classes. A majority of the models performed well in distinguishing between normal and viral pneumonia chest X-ray images. Most of the models had accuracy scores of upto 99% and F1-Score values of over 92% in all three Transfer learning techniques used. The second-best classification category was Normal vs COVID-19, followed by normal vs viral pneumonia vs COVID-19, and the worst performing classification category was normal vs viral pneumonia vs COVID-19 vs lung opacity. The steady decrease in model performance with an increase in the number of dataset classes can be linked to an increase in the number of data specific features that a model has to learn. Hence the model may be prone to misprediction when some patterns across the dataset classes overlap. Moreover, the time taken to train the CNNs in multi-class classification is more than the time taken to train them in binary classification mainly because of an increase in the size of the training data. There was also minimal to no overfitting experienced when training models using all the three Transfer learning techniques. This can be attributed to the robust regularization techniques used such as data augmentation, dropout and L2 regularization.

The EfficientV2B2, MobileNetV3Large and MobileNetV3Small are the three top performing models in all the experiments undertaken in the study. The three models generated consistently high results in all the three Transfer learning techniques, that is, feature extraction transfer learning with SVM and dense layer classifier, and Fine-tuning transfer learning. However, EfficientNetV2B2 CNN had the second longest training time after the Xception neural net. MobileNetV3Small and MobileNetV3Large were the most suitable models for the study. Not only did the models give excellent model performance metric values, but also the two models also had the shortest training time in all the conducted experiments. Xception neural net was the most unsuitable model because it had the longest training time and was also consistently amongst the two worst performing models in all the conducted experiments.

The study results indicate that deep transfer learning can be considered as an effective and efficient alternative to creating complex model architectures from scratch. The pre-trained models can produce better outcomes using minimal time and effort. The exceptional outcomes from this research work also suggest that Transfer learning can be used in the development of AI tools for automatic detection of COVID-19, non-COVID-19 viral pneumonia and other lung infections. Moreover, viral pneumonia and COVID-19 both being viral diseases exhibit similar signs and symptoms and can easily be misdiagnosed. Hence, the use of AI disease detection tools can help prevent cases of misdiagnosis and ensure proper treatment is administered to patients. The AI detection tools can also help in early screening and identification of COVID-19 infected persons which can help prevent the rapid spread of the disease.

Limited availability of training dataset was one of the limitations in the study, hence the use of image augmentation to generate more training dataset and ensure the dataset was balanced. In future research work, the number of original training dataset can be increased when training the models. Furthermore, fine-tuning offers limitless ways of tuning model parameters. Hence, more hyper-parameter tuning can be conducted to further boost the models’ performance.

REFERENCES

Basha, S. S., Dubey, S. R., Pulabaigari, V., & Mukherjee, S. (2020). Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, *378*, 112–119. https://doi.org/10.1016/j.neucom.2019.10.008

Choi, Y. A., Park, S. J., Jun, J. A., Pyo, C. S., Cho, K. H., Lee, H. S., & Yu, J. H. (2021). Deep Learning-Based Stroke Disease Prediction System Using Real-Time Bio Signals. *Sensors*, *21*(13), 4269. https://doi.org/10.3390/s21134269

Dahmani, M., Chowdhury, M. E. H., Khandakar, A., Rahman, T., Al-Jayyousi, K., Hefny, A., & Kiranyaz, S. (2020). An Intelligent and Low-Cost Eye-Tracking System for Motorized Wheelchair Control. *Sensors*, *20*(14), 3936. https://doi.org/10.3390/s20143936

Day, O., & Khoshgoftaar, T. M. (2017). A survey on heterogeneous transfer learning. *Journal of Big Data*, *4*(1). https://doi.org/10.1186/s40537-017-0089-0

Gao, X. W., & QIan, Y. (Eds.). (2017). . *Application of deep learning neural network for classification of TB lung CT images based on patches*. CEUR Workshop Proceedings.

Guérin, J., Thiery, S., Nyiri, E., Gibaru, O., & Boots, B. (2021). Combining pretrained CNN feature extractors to enhance clustering of complex natural images. *Neurocomputing*, *423*, 551–571. https://doi.org/10.1016/j.neucom.2020.10.068

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187,* 27–48. https://doi.org/10.1016/j.neucom.2015.09.116

Hahn, S., & Choi, H. (2020, July). Understanding dropout as an optimization trick. *Neurocomputing*, 64–70. https://doi.org/10.1016/j.neucom.2020.02.067

Haleem, A., Javaid, M., & Vaishya, R. (2020). Effects of COVID-19 pandemic in daily life. *Current Medicine Research and Practice, 10*(2), 78–79. https://doi.org/10.1016/j.cmrp.2020.03.011

Irani, S. (2022). Immune responses in SARS-CoV-2, SARS-CoV, and MERS-CoV infections: A comparative review. *International Journal of Preventive Medicine*, *13*(1), 45. https://doi.org/10.4103/ijpvm.ijpvm_429_20

Jo, T., Nho, K., & Saykin, A. J. (2019). Deep Learning in Alzheimer’s Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, *11*. https://doi.org/10.3389/fnagi.2019.00220

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*, *74*(5), 338–345. https://doi.org/10.1016/j.crad.2018.12.015

Karoly, A. I., Galambos, P., Kuti, J., & Rudas, I. J. (2021). Deep Learning in Robotics: Survey on Model Structures and Training Strategies. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *51*(1), 266–279. https://doi.org/10.1109/tsmc.2020.3018325

Krittanawong, C., Johnson, K. W., Rosenson, R. S., Wang, Z., Aydar, M., Baber, U., Min, J. K., Tang, W. H. W., Halperin, J. L., & Narayan, S. M. (2019). Deep learning for cardiovascular medicine: a practical primer. *European Heart Journal*, *40*(25), 2058–2073. https://doi.org/10.1093/eurheartj/ehz056

Kumar, A., Abhishek, K., Ghalib, M. R., Nerurkar, P., Shah, K., Chandane, M., Bhirud, S., Patel, D., & Busnel, Y. (2020). Towards cough sound analysis using the Internet of things and deep learning for pulmonary disease prediction. *Transactions on Emerging Telecommunications Technologies*. https://doi.org/10.1002/ett.4184

Lee, C. Y., Gallagher, P., & Tu, Z. (2018, April). Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 863–875. https://doi.org/10.1109/tpami.2017.2703082

Lee, G., Park, H., Kim, N., Yu, J., Jo, S., & Choi, K. (2019). Acceleration of DNN Backward Propagation by Selective Computation of Gradients. *Proceedings of the 56th Annual Design Automation Conference 2019*. https://doi.org/10.1145/3316781.3317755

Lemley, J., Bazrafkan, S., & Corcoran, P. (2017). Smart Augmentation Learning an Optimal Data Augmentation Strategy. *IEEE Access*, *5*, 5858–5869. https://doi.org/10.1109/access.2017.2696121

Li, Q. (2020). An Outbreak of NCIP (2019-nCoV) Infection in China — Wuhan, Hubei Province, 2019−2020. *China CDC Weekly*, *2*(5), 79–80. https://doi.org/10.46234/ccdcw2020.022

Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., & Zhang, G. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, *80*, 14–23. https://doi.org/10.1016/j.knosys.2015.01.010

Internet: Mahmud, F. (2020, March 24). Bangladesh scientists create $3 kit. Can it help detect COVID-19? Web: https://www.aljazeera.com/news/2020/3/24/bangladesh-scientists-create-3-kit-can-it-help-detect-covid-19. Last accessed date: 15.07.2022

Mambou, S., Maresova, P., Krejcar, O., Selamat, A., & Kuca, K. (2018). Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model. *Sensors*, *18*(9), 2799. https://doi.org/10.3390/s18092799

Marques, G., Agarwal, D., & de la Torre Díez, I. (2020). *Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network* (Vol. 96). Elsevier BV. https://doi.org/10.1016/j.asoc.2020.106691

Mehtar, S., Preiser, W., Lakhe, N. A., Bousso, A., TamFum, J. J. M., Kallay, O., Seydi, M., Zumla, A., & Nachega, J. B. (2020). Limiting the spread of COVID-19 in Africa: one size mitigation strategies do not fit all countries. *The Lancet Global Health*, *8*(7), e881–e883. https://doi.org/10.1016/s2214-109x(20)30212-6

Michelucci, U. (2018). *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*. Apress.

Mutasa, S., Sun, S., & Ha, R. (2020). Understanding artificial intelligence based radiology studies: What is overfitting? *Clinical Imaging*, *65*, 96–99. https://doi.org/10.1016/j.clinimag.2020.04.025

Niu, S., Liu, Y., Wang, J., & Song, H. (2020). A Decade Survey of Transfer Learning (2010–2020). *IEEE Transactions on Artificial Intelligence*, *1*(2), 151–166. https://doi.org/10.1109/tai.2021.3054609

Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2018). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, *32*(15), 10927–10933. https://doi.org/10.1007/s00521-018-3689-5

Panesar, A. (2019). *Machine Learning and AI for Healthcare* (1st ed.). Apress.

Peirelinck, T., Kazmi, H., Mbuwir, B. V., Hermans, C., Spiessens, F., Suykens, J., & Deconinck, G. (2022). Transfer learning in demand response: A review of algorithms for data-efficient modelling and control. *Energy and AI*, *7*, 100126. https://doi.org/10.1016/j.egyai.2021.100126

Rahaman, M. S. (2020). Home Quarantine Guidelines: COVID 19 Panorama. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3575349

Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S. B., Islam, M. T., al Maadeed, S., Zughaier, S. M., Khan, M. S., & Chowdhury, M. E. (2021). Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in Biology and Medicine*, *132*, 104319. https://doi.org/10.1016/j.compbiomed.2021.104319

Rajaraman, S., & Antani, S. (2020). Weakly Labeled Data Augmentation for Deep Learning: A Study on COVID-19 Detection in Chest X-Rays. *Diagnostics*, *10*(6), 358. https://doi.org/10.3390/diagnostics10060358

Ruuskanen, O., Lahti, E., Jennings, L. C., & Murdoch, D. R. (2011). Viral pneumonia. *The Lancet*, *377*(9773), 1264–1275. https://doi.org/10.1016/s0140-6736(10)61459-6

Saidi, A., Othman, S. B., & Saoud, S. B. (2020, December). Hybrid CNN-SVM classifier for efficient depression detection system. *2020 4th International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*. https://doi.org/10.1109/ic_aset49463.2020.9318302

Sajjad, O., Shahzad, A., & Mahmood, S. (2020). RT-PCR for COVID-19: Diagnosis Made Easy. *BioMedica*, *36*(2S), 115–120. https://doi.org/10.51441/biomedica//biomedica/5-451

Shekar, G., Revathy, S., & Goud, E. K. (2020). Malaria Detection using Deep Learning. *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. https://doi.org/10.1109/icoei48184.2020.9143023

Tilve, A., Nayak, S., Vernekar, S., Turi, D., Shetgaonkar, P. R., & Aswale, S. (2020). Pneumonia Detection Using Deep Learning Approaches. *2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE)*. https://doi.org/10.1109/ic-etite47903.2020.152

Tran, P. T., & Phong, L. T. (2019). On the Convergence Proof of AMSGrad and a New Version. *IEEE Access*, *7*, 61706–61716. https://doi.org/10.1109/access.2019.2916341

Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., Doel, T., David, A. L., Deprest, J., Ourselin, S., & Vercauteren, T. (2018). Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging*, *37*(7), 1562–1573. https://doi.org/10.1109/tmi.2018.2791721

Wiatowski, T., & Bolcskei, H. (2018). A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction. *IEEE Transactions on Information Theory*, *64*(3), 1845–1866. https://doi.org/10.1109/tit.2017.2776228

World Health Organization. (2020, March). *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

Zhang, C., Zhao, Q., & Wu, H. (2022). Deep domain adaptation via joint transfer networks. *Neurocomputing*, *489*, 441–448. https://doi.org/10.1016/j.neucom.2022.03.028

Zhang, J., Li, Y., Tian, J., & Li, T. (2018). LSTM-CNN Hybrid Model for Text Classification. *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. https://doi.org/10.1109/iaeac.2018.8577620

Zhang, L., Li, Y., Zhou, J., Li, J., Tong, C., Cai, J., Yang, Z., Li, S., Zhu, J., Zhao, M., Bai, Z., Xie, X., & Chen, T. (2021). Comparison of the community-acquired pneumonia and COVID-19 at the early stage: findings from two cohort studies. *Annals of Palliative Medicine*, *10*(9), 9572–9582. https://doi.org/10.21037/apm-21-2006

Zhou, K., Greenspan, H., & Shen, D. (2017). *Deep Learning for Medical Image Analysis (The MICCAI Society book Series)* (1st ed.). Academic Press.

GAZİLİ OLMAK AYRICALIKTIR...