Tagging Hadronically Decaying Top Quarks with Deep Neural Networks

Author:
Elise HINKLE

Supervisor:
Dr. Meenakshi NARAIN
Bjorn BURKLE

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Science in Mathematical Physics in the

Department of Physics
Brown University

April 30, 2019

BROWN UNIVERSITY

Abstract

Meenakshi Narain
Department of Physics

Bachelor of Science

Tagging Hadronically Decaying Top Quarks with Deep Neural Networks

by Elise HINKLE

The top quark is one of the most important Standard Model particles in probing new physics beyond the Standard Model. Experimentally, this exploration can be done by studying proton-proton collisions in the Large Hadron Collider with the Compact Muon Solenoid (CMS) detector. There must, however, be a reliable method to identify top quarks or their signatures in CMS data. As the top quark events of greatest interest are often Lorentz boosted, they appear as jets in detector data and can be tagged using jet substructure techniques. Increasingly, these top tagging techniques involve the use of deep learning tools such as Deep Neural Networks (DNNs). In this study, we compare the top-tagging performance of two CMS top-tagging neural networks, the Boosted Event Shape Tagger (BEST) and DeepAK8, as well as a new DNN which takes as inputs the jet topological and kinematic BEST input variables and the particle-level features used as inputs to DeepAK8. We find that this new network, PF+BEST, performs comparably to the two established networks with significant promise for improvement and utility in future new physics searches involving coupling to top quarks.

Acknowledgements

I would like to express my deepest thanks and appreciation to all of the people who assisted and supported me throughout this project. First and foremost, my advisor, Professor Meenakshi Narain, and Bjorn Burkle, the graduate student also working on my project, have taught me so much about top tagging, machine learning, and jet substructure based tagging techniques in general. They have also been amazing mentors outside of this particular project and have given me advice and knowledge I will hold onto long after I leave Brown. Furthermore, I want to thank current and former members of the Brown CMS Group including Ginger Cheng, Mary Hadley, John Hakala, Marlene Ortega, Eric Scotti, Rizki Suarif, Vuong Truong, Emanuele Usai, and Jess Wong for their help and support on this project and throughout my time in the CMS group at Brown. This project and report would not have been possible without all of you. In particular, I greatly appreciate Dr. Suarif's help processing MiniAOD files with LJMet, Dr. Usai sharing his knowledge of jet tagging and machine learning, and Dr. Usai's and Mary Hadley's valuable feedback on this paper and its corresponding presentation. I also want to thank Dr. Julie Hogan of Bethel University for all of her assistance understanding LJMet, BEST, and DeepAK8.

In addition to all of the members of my lab group, I want to thank all of my friends and family for their support through this project and throughout my time at Brown. I particularly want to thank Jason Chan, Jamie Holber, Rahul Jayaraman, Hope McGovern, and David Paasche for all of the hours we spent together working on lab reports, problem sets, and studying for physics exams. I also want to thank all current and past players and staff of Brown Women's Rugby, in particular, my coach, Kathy Flores, for their support of and belief in me over the past four years. Finally, I would like to thank my parents, Bethanne and Chris, and my siblings, Trevor and Adele, for all of their love and support for the past twenty-two years.

Contents

List of Figures

List of Tables

List of Abbreviations

ANN	**A**rtificial **N**eural **N**etwork
ATLAS	**A** **T**oroidal **L**HC **A**pparatu**S**
AUC	**A**rea **U**nder (the ROC) **C**urve
BEST	**B**oosted **E**vent **S**hape **T**agger
BSM	**B**eyond (the) **S**tandard **M**odel
CERN	**C**onseil **E**uropéen (pour la) **R**echerche **N**ucléaire
CMS	**C**ompact **M**uon **S**olenoid
CNN	**C**onvolutional **N**eural **N**etwork
DNN	**D**eep **N**eural **N**etwork
ECAL	(Crystal) **E**lectromagnetic **CAL**orimeter
EDM	**E**vent **D**ata **M**odel
EWSB	Electro**W**eak **S**ymmetry **B**reaking
GPU	**G**raphics **P**rocessing **U**nit
HCAL	**H**adron **CAL**orimeter
HDF5	**H**ierarchical **D**ata **F**ormat (**5**)
HL-LHC	**H**igh-**L**uminosity **LHC**
LHC	**L**arge **H**adron **C**ollider
MC	**M**onte **C**arlo
MiniAOD	**Mini**ature **A**dvanced **O**ptical **D**isc
ML	**M**achine **L**earning
PF	**P**article **F**low
QCD	**Q**uantum **C**hromo**D**ynamics
QFT	**Q**uantum **F**ield **T**heory
ROC	**R**eceiver **O**perating **C**haracteristic
SM	**S**tandard **M**odel

List of Symbols

$m_{t\bar{t}}$	invariant top pair mass	eV
p_T	transverse momentum	eV
R	jet radius parameter	
$\sqrt{s}$	center of mass energy	eV
η	pseudorapidity	
ϕ	azimuthal scattering angle	rad
τ_{21}	two-pronged N-subjettiness	
τ_{32}	three-pronged N-subjettiness	

Chapter 1

Introduction & Background

1.1 The Standard Model

In the field of elementary particle physics, the fundamental building blocks of the universe and their interactions are described through the Standard Model (SM) of particle physics. According to the SM, there are three forces acting on these subatomic particles, the strong force, the weak force, and the electromagnetic force, with each force propagated by at least one spin-1 vector boson: the strong force is propagated by gluons (g), the weak force is propagated by W and Z bosons (W^+ , W^-, and Z), and the electromagnetic force is propagated by photons (γ). The fermions, or spin-$\frac{1}{2}$ particles, predicted by the SM include three generations of charged leptons (electron, e, muon, μ, and tau lepton, τ), which interact through the electromagnetic and weak forces, three corresponding generations of neutrinos, or uncharged leptons (ν_e, ν_μ, and ν_τ), which only interact through the weak force, and three generations each of down-type (down, d, strange, s, and bottom, b) and up-type (up, u, charm, c, and top, t) quarks, which interact through the strong, electromagnetic, and weak forces. Each fermion also has a corresponding antimatter particle, or antiparticle, with opposite charge [1]. An antiparticle is denoted by a bar; for instance, a top antiquark is denoted as $\bar{t}$.

Until the late 2000s, these particles and forces composed the observable aspects of the Standard Model. In 2012, however, two experiments situated on the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) found evidence for an addition to the SM [2, 3]. The two experiments, A Large Toroidal LHC Apparatus (ATLAS) and Compact Muon Solenoid (CMS), identified this signature as the Higgs boson, a particle theorized to be included in the SM but not observed before that point. The Higgs boson, unlike W and Z bosons, is a scalar boson and is significant because its interactions with other elementary SM particles give them mass [2, 3]. A general overview of the SM, including the Higgs boson, can be seen in Figure 1.1.

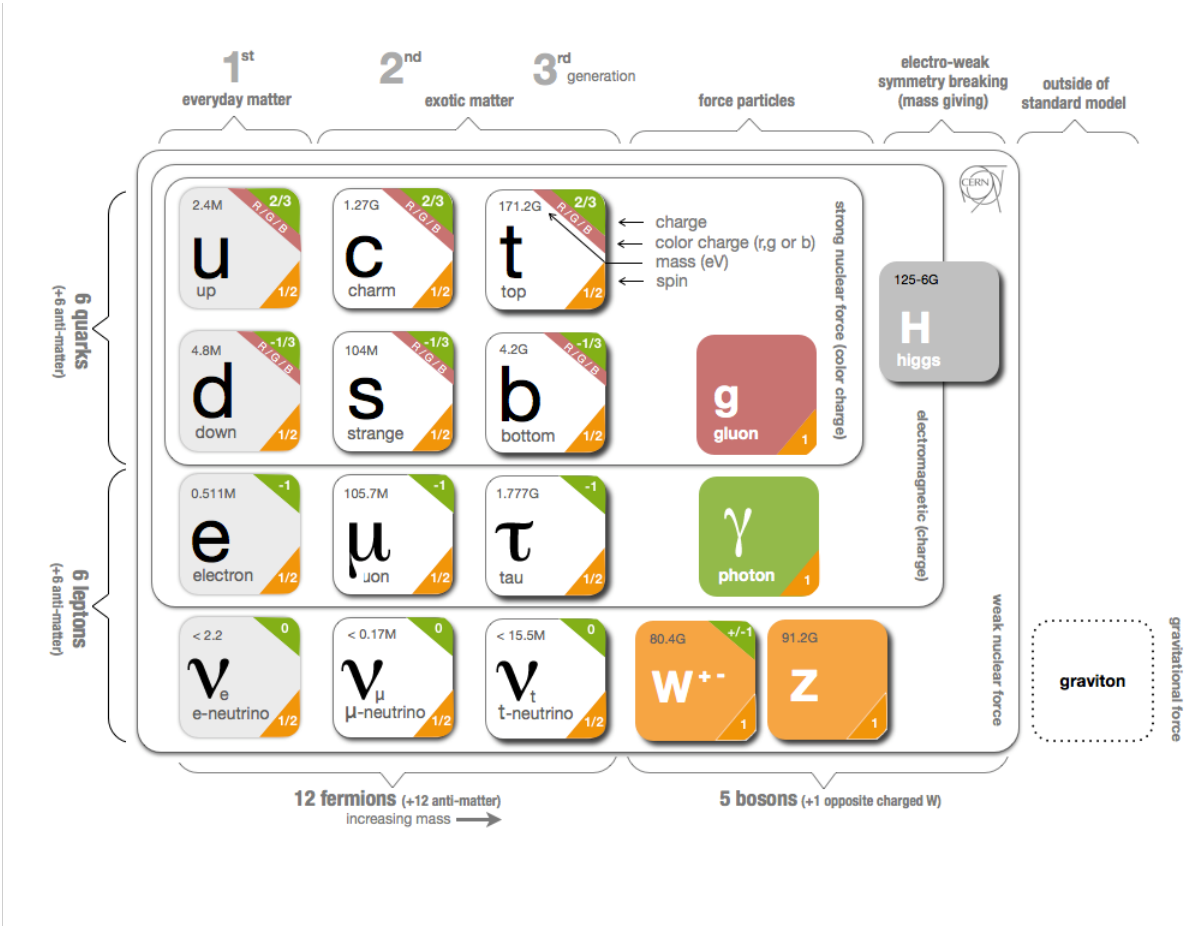


FIGURE 1.1: An overview of SM particles, force carriers, and forces. Also pictured is the graviton, a theorized carrier of gravitational force which currently exists outside of the SM [4].

1.2 Physics Beyond the Standard Model and Top Quarks

While the discovery of the Higgs boson exhibited some sense of ongoing consistency between SM theoretical and observable physics, there are still inconsistencies between SM theory and observation. These inconsistencies range from charge-parity violation in certain interactions, which leads to the asymmetry of matter and anti-matter in the universe, to the exclusion of gravity from the SM and the lack of an explanation for so-called "dark matter" and "dark energy," entities which are evident from their cosmological effects but not registered in SM theory or generally well-understood. Such physics topics are known as physics Beyond the SM, or BSM physics.

There are many models of BSM physics yet to be effectively interrogated by experimentalists. A number of these models can be constrained or supported by evidence for top quark production and decay. For instance, even though a Higgs boson had been theorized before its observation in 2012, its measured mass of 125.1 GeV is light – in comparison, the top quark is 173.1 GeV [3, 5]. The unnaturalness of observed particle mass hierarchy leads to possibilities of new physics particles through a variety of theories. This so-called hierarchy problem could, for instance, be solved by quantum field theory (QFT)-based corrections to the mass of the Higgs boson [6]. These corrections all involve masses heavier than the Higgs itself and arise from couplings of the Higgs boson to $W^{\pm}$ and Z bosons, photons, the Higgs boson, and top quarks [6]. Such corrections, known as loop corrections, can be seen in Figure 1.2.

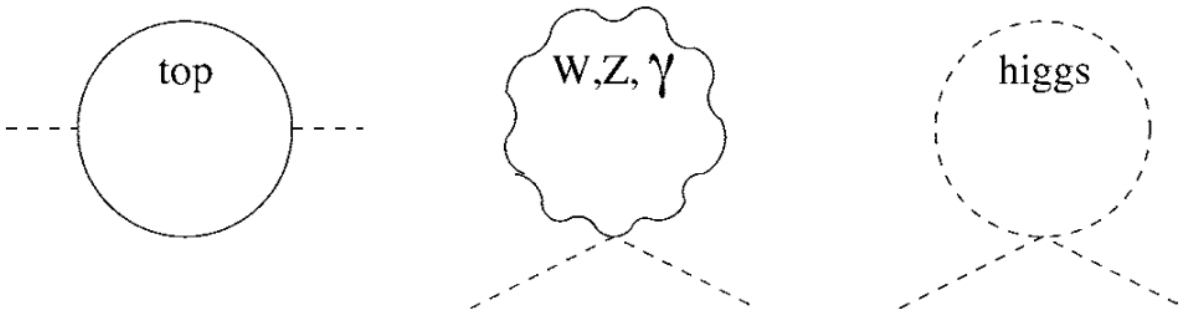


FIGURE 1.2: Feynman diagrams showing the largest corrections to the Higgs mass. These corrections arise from QFT-based "loop" corrections involving top quarks, electroweak gauge bosons, and the Higgs boson itself [6].

Further suggestions for new physics models include those which suggest additional sources of Electroweak Symmetry Breaking (EWSB), such as composite Higgs models and the topcolor model with a Z' boson, and involve new heavy particles which interact strongly with top quarks [7]. EWSB mechanisms are responsible for the fact that the electromagnetic force carrier, the photon, is massless while weak force carriers, $W^{\pm}$ and Z bosons, are massive particles. As SM sources of EWSB do not completely explain experimental results, it is important to probe additional methods of EWSB [8]. Other BSM physics models which can be better understood by studying top quarks include the Randall-Sundrum Kaluza-Klein model, which describes a new massive gluon which decays to top quarks, and models such as two-Higgs-doublet models which extend the SM Higgs doublet model [7]. These models, as well as the models implicated in EWSB mechanism searches described above, also all provide solutions to the hierarchy problem [7].

1.3 LHC Physics and CMS

While it is helpful to understand the theories which could explain the inconsistencies between the SM and current observations, continuing observation and experimentation is necessary in order to further constrain or support existing theories and to model independent measurements which could guide future theories. The two most significant experiments which can make the necessary measurements are the ATLAS and CMS experiments on the LHC. Both ATLAS and CMS have the ability to detect leptons, hadrons, and photons resulting from the proton-proton ($p\bar{p}$) collisions in the LHC. The two detectors are, however, designed slightly differently, with the CMS detector able to achieve a higher energy resolution for electrons and photons and the ATLAS detector able to achieve better charged particle momentum resolution [7]. Thus, each experiment has its advantages and disadvantages. Their dual, yet separate, existence allows each collaboration to develop unique experimental techniques and procedures to understand observable data and make independent discoveries. The research reported in this paper is focused on data and techniques specific to the CMS Collaboration.

Chapter 2

The CMS Detector and Jet Detection

2.1 CMS Detector Overview

The CMS detector is located at one of the four $p\bar{p}$ collision points in the 27-km-around LHC [9]. The detector is 15 meters high, 21 meters long, and weighs 14,000 tonnes. Its detector components are arranged radially around the detector in layers and include silicon trackers, a superconducting solenoid magnet, a crystal electromagnetic calorimeter (ECAL), a scintillating hadron calorimeter (HCAL), and muon drift chambers. A diagram of the detector can be seen in Figure 2.1.

Each layer plays a different and important role in detecting particles resulting from $p\bar{p}$ collisions. The superconducting solenoid magnet, from which CMS gets part of its name, creates a magnetic field of 3.8 Tesla which bends the path of charged particles. As particles bend in different directions based on the sign of their charge and lower momentum particles bend less than their higher momentum counterparts, the solenoid's bending of charged particle trajectories assists in measuring particle charge and momentum. The inner-most layer of silicon trackers is used to reconstruct the trajectory of the charged particles bent by the solenoid magnet's field. Next, the ECAL and HCAL measure the energies of electrons and photons, and hadrons respectively. Finally, muons, which are not stopped by either calorimeter, travel through the outermost layer of muon chambers, which measure the muon's momentum a second time (the initial measurement is done by the silicon trackers) [9]. A cross section of the detector, including the way in which different particles might interact with the detector, can be seen in Figure 2.2 [10].

It is also important to note that neutrinos, which interact solely through the weak force, can only be "observed" in the CMS detector as missing transverse energy. It is, however, impossible to definitively identify missing energy as a neutrino or neutrinos because there are many theories which involve some sort of new particle which might appear as missing energy or which might itself decay to neutrinos. Thus, any missing energy found in events expected to include neutrinos could either be neutrinos from a previously studied interaction or related to new physics.

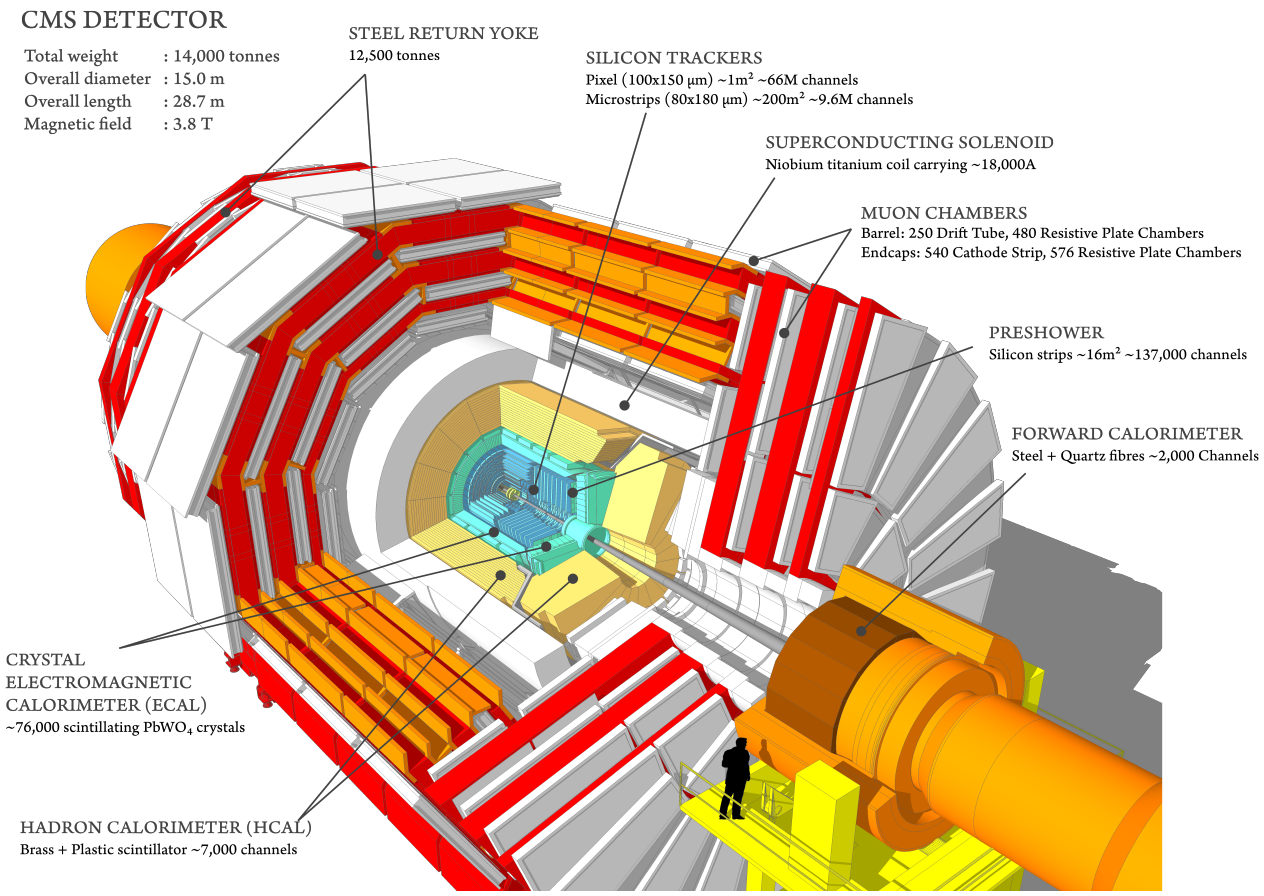


FIGURE 2.1: An artistic representation of the CMS Detector and all of its significant components. The detector is centered around the LHC beam line such that most particles travelling outward from a $p\bar{p}$ collision in the portion of the LHC running through the center of the detector move radially outward through the detector. A person is seen standing on the yellow platform for scale [9].

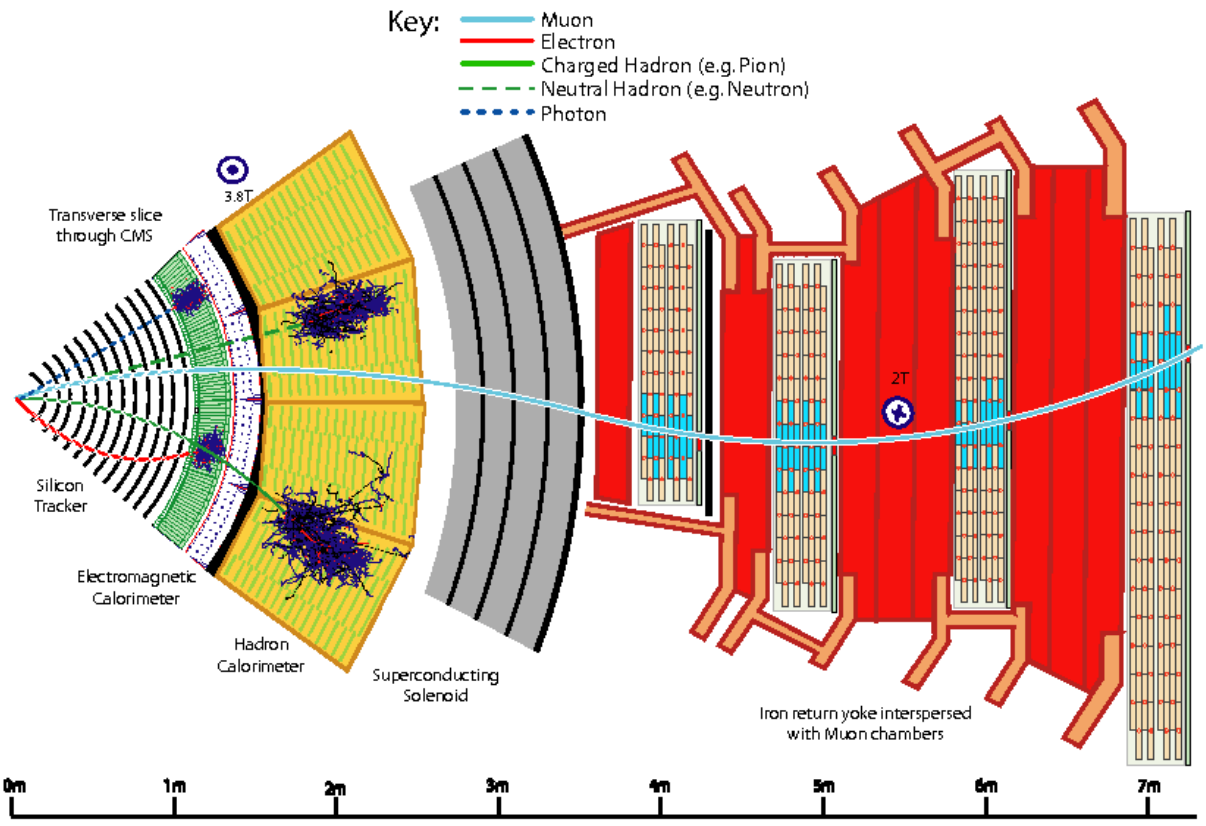


FIGURE 2.2: A "slice" of the CMS detector showing each of the layers and how five different classes of particles might move through the detector [10].

Chapter 3

Identifying Top Quarks with CMS

3.1 $t\bar{t}$ Production and Decay

In order to use top quarks to investigate different BSM physics models, it is first important to be able to identify top quarks in detector data. A simple reconstruction of top quarks from decay particles is difficult due to background quantum chromodynamics (QCD) interactions, which have signatures similar to top quark "signal" events [11]. Top quark production and decay in $p\bar{p}$ collisions might be described by one of the following interactions:

$$q\bar{q} \rightarrow t\bar{t} \rightarrow bW^{+}\bar{b}W^{-} \tag{3.1}$$

$$gg \rightarrow t\bar{t} \rightarrow bW^{+}\bar{b}W^{-} \tag{3.2}$$

At the LHC, top pairs are most often produced through gluonic interactions. The W^{+} and W^{-} in this decay can either decay leptonically to a charged lepton and a neutrino or hadronically to two quarks. Thus, the entire final state consists of two b quarks and four other quarks (fully hadronic decay); two b quarks, two other quarks, one charged lepton, and one neutrino (semi-leptonic decay); and two b quarks, two charged leptons, and two neutrinos (fully leptonic decay) [1]. A Feynman diagram of the semi-leptonic decay channel can be seen in Figure 3.1.

Unfortunately, neutrinos cannot be definitively identified in CMS, as described in Chapter 2.1. Thus, it is impossible to fully reconstruct top pairs ($t\bar{t}$) which decay leptonically or semi-leptonically. Due to this limitation and its implications in particular for new physics searches, we chose to only study $t\bar{t}$ events which decay hadronically to two b quarks and 4 other quarks [11].

In studying such hadronic $t\bar{t}$ decays in the CMS detector, it is important to note that all final state quarks, including the b quarks seen in Equations 3.1 and 3.2, are seen in the detector as "jets" instead of distinct particles. This effect is due to gluonic interactions in QCD described as color confinement which prohibit the observation of singular quarks or gluons. Instead, quarks and gluons appear as hadrons, or bound states of multiple quarks

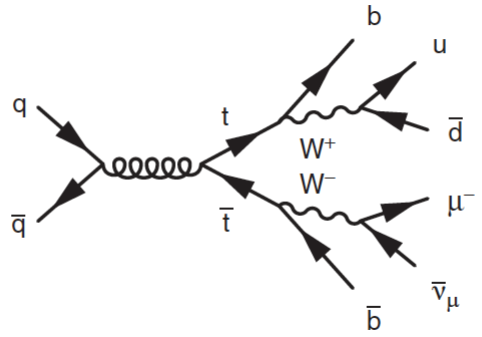


FIGURE 3.1: A Feynman diagram of the lowest order showing production of $t\bar{t}$ event from a $p\bar{p}$ collision such as those in the LHC. As one W boson decays to two leptons and the other decays to two quarks, this decay, as classified by its final state, is semi-leptonic [1].

and/or gluons. In high energy collisions such as the $p\bar{p}$ collisions in the LHC, hadrons are observed as a circular collection of energy signatures and, in the case of charged hadrons, charge signatures in the detector which can be traced back to the primary interaction vertex, forming a cone. This cone of particles arising from the hadronization of a quark is known as a jet [7]. Thus, hadronic top decays in the CMS detector are observed as two b jets and four other jets [1].

3.2 Boosted Top Decays

Hadronic top decays at low transverse momentum (p_T) can be modeled as described above as two b jets and four jets from the decay of the W boson [1]. For top decays related to couplings with heavy new physics particles, the p_T of the top decays are high enough that the event becomes Lorentz boosted, effectively collimating the decay and allowing the top decay to present as a jet itself [1, 12]. A model of this process in different p_T regimes can be seen in Figure 3.2.

3.3 CMS Jet Reconstruction

There are three main layers of the CMS detector primarily concerned with the detection and tagging of jets: the HCAL, the ECAL, and the silicon trackers [10]. The energy clusters measured by the HCAL and ECAL are used to reconstruct jets [7]. This data can be combined with charged particle trajectory information from the silicon trackers in order to provide a rich reconstruction of the substructures of jets. In general, measurements

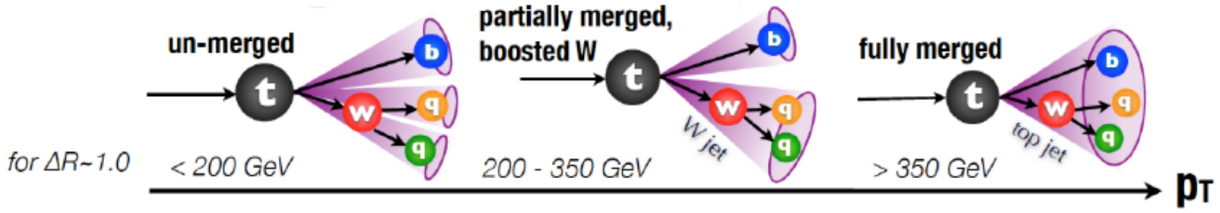


FIGURE 3.2: A schematic showing the way in which the subjets of a top quark decay merge with increasing p_T. The $\Delta R \sim 1.0$ note refers to the jet angular distance parameter. This schematic was created for LHC $p\bar{p}$ collisions with center of mass energy ($\sqrt{s}$) of 13 TeV [12].

from all detector layers can be combined and correlated in order to reconstruct final state particle properties as part of a technique known as particle-flow (PF) reconstruction. PF reconstruction allows physicists analyzing CMS data to better understand and utilize the substructure of jets [10]. In this case, jets are clustered and reconstructed with the anti-k_t algorithm using a jet radius parameter R of 0.4 for so-called AK4 jets and a jet radius parameter R of 0.8 for so-called AK8 jets [13, 14].

3.4 Top Tagging Through Jet Substructure

As many of the BSM physics searches are looking in regimes of high top pair invariant masses ($m_{t\bar{t}}$) over 1 TeV, the resulting top decays are highly boosted and must be observed as top jets. Thus, instead of looking for the jets of each separate top decay product, it is necessary to probe the substructure of the boosted jet in order to identify it as a top jet and, subsequently, a decaying top quark [7]. Specifically, jet substructure identification techniques involve analyzing the subjets, or jet constituents, which make up the "main" jet. By only considering subjets in particular ranges of kinematic variables, such as jet p_T and invariant mass, the top tagging algorithm can ideally distinguish between a "true" top jet and a background event. A background event is, in this case, a non-top jet which could be mistaken for a top jet. The main source of background for top jet tagging comes from QCD multijet events. In addition to top jet classification, these jet substructure methods are used for classifying signatures of boosted W, Z, and Higgs bosons [15].

As the CMS collaboration can simulate large numbers of realistic jet events which can be labeled by their decay particle and described by the observable jet substructure features, machine learning (ML) algorithms are commonly applied to jet substructure-based tagging algorithms. These simulations allow for the use of ML techniques because they can act as training and testing data sets so that a computer can "learn" how to identify a top jet, for instance, in detector data. Simple neural networks, decision trees, and random forests are examples of initial ML algorithms used in tagging algorithms. While

these initial taggers might have relied on simple substructure variables such as the number of subjets and the invariant jet mass, physicists are now able to take advantage of more advanced ML architectures. For instance, artificial neural networks (ANNs) and, in particular, deep neural networks (DNNs) with multiple layers and inputs allow physicists to take greater advantage of the richness of a jet's substructure [16]. These taggers range from those which have image-like jet representations as inputs to networks similar to those used for traditional image classification, to networks which use jet constituent variable-ordered representations of jets as inputs to networks similar to those used for natural language processing [16].

3.5 Exploring ANN- and DNN-based Top Tagging in CMS

While there are many ways to use ANNs and DNNs to classify jets by their substructure, it is important to better understand which inputs to these algorithms lead to the most effective classification and which jet substructure variables can be learned by the network. It is also important to recognize when a technique is taken too far and the true physics is overlooked [16]. In this study, we compared networks using the inputs of two CMS jet substructure-based tagging algorithms, the Boosted Event Shape Tagger (BEST), an ANN, and DeepAK8, a DNN, as well as a new, third DNN tagger which trains on the inputs of both BEST and DeepAK8, in the hopes of better understanding the classes and combinations of inputs which allow for the most effective top tagger.

3.6 BEST

BEST is a boosted jet classification algorithm which employs a neural network. In general, the algorithm boosts values derived from measurements of individual jet constituent particles into the rest decay frames of different massive particles. Heavy particles considered include $W^{\pm}$, Z, and Higgs bosons, and top quarks. As the masses of these particles have been measured experimentally ($m_t = 173.1$ GeV, $m_W = 80.4$ GeV, $m_Z = 91.2$ GeV, and $m_t = 125.1$ GeV), their four-vector $p_\mu \equiv (\vec{p}, E)$ can be easily computed. For instance, the four-vector of the top quark can be computed as:

$$p_\mu^{(t)} = (\vec{p}, m_t^2 + p^2). \tag{3.3}$$

Such four vectors can therefore be used to find the boost vectors needed to transform the jet constituent values from the boosted detector frame to a rest decay frame. By subsequently comparing the jet constituent values with the rest frames of these particles, the

tagger classifies a jet as originating from one of these particles or a background light-flavor QCD jet. Thus, BEST can be useful in tagging a variety of heavy objects in addition to boosted top jets [5, 17].

3.6.1 BEST Input Variables

The neural network which supports the BEST classifier takes in 59 separate jet constituent observable quantities. All variables are listed in Table 3.1. The variables in the middle three columns are boosted into each hypothesis rest frame, while all other variables are based on the original detector-frame jet value [5]. Jet η, or pseudorapidity, is a value derived from the polar angle of the jet in relation to the beam axis, with higher η values indicating a smaller angle between jet and beam axis [1]. Jet τ_{21} and τ_{32} are values of N-subjettiness, or the likelihood that a decay is two-pronged (i.e. for, $W^{\pm}$, Z, and Higgs bosons) or three-pronged (i.e. for top quarks) respectively [18]. The jet charge is found from summing the charges of all jet constituents as weighted by p_T. Sphericity and aplanarity consider the momentum distribution of the jet constituents, isotropy measures the uniformity of the jet constituents' spatial distribution, and thrust is related to the maximum p_T along a certain axis. The angular distribution of a group of jet constituents is encoded in the Fox-Wolfram Moment variables, and longitudinal asymmetry contains information about the vector boosting the particles to a given rest frame. The masses listed in the fourth column come from reclustering the jet constituents after boosting them to a rested frame and considering the masses of different combinations of the first through fourth leading jets [5]. The subjet CSV Values are from a tagging algorithm used to identify bottom quark subjets and the soft-drop jet mass is the mass of a jet after removing mass derived from wide-angle radiation [19, 20]. For more information on the BEST input observables, please refer to CMS AN-18-095 or the BEST group's 2016 paper in Physical Review D [5, 17].

3.6.2 BEST Neural Network Architecture

The version of BEST used in this paper had three 40-node, hidden layers and used a rectified-linear activation function. It was trained with the scikit-learn Python package using the MLPClassifier module on 500,000 events from samples of simulated AK8 jets split between 6 training samples. The samples used were based on CMS Run II data and were made in Spring 2016 and had a jet p_T range between 500 GeV and 2 TeV. For more information on the BEST neural network, its training, and its performance please refer to CMS AN-18-095 or the BEST group's 2016 paper in Physical Review D [5, 17].

BEST Input Variables				
Jet η	Sphericity (t,W,Z,H)	Fox-Wolfram Moment H_1/H_0 (t,W,Z,H)	m_{12} (t,W,Z,H)	Subjet 1 CSV Value
Jet τ_{21}	Aplanarity (t,W,Z,H)	Fox-Wolfram Moment H_2/H_0 (t,W,Z,H)	m_{23} (t,W,Z,H)	Subjet 2 CSV Value
Jet τ_{32}	Isotropy (t,W,Z,H)	Fox-Wolfram Moment H_3/H_0 (t,W,Z,H)	m_{13} (t,W,Z,H)	Max Subjet CSV Value
Jet Charge	Thrust (t,W,Z,H)	Fox-Wolfram Moment H_4/H_0 (t,W,Z,H)	m_{1234} (t,W,Z,H)	Jet soft-drop mass
		Longitudinal Asymmetry (t,W,Z,H)		

TABLE 3.1: Table showing the input variables to the BEST algorithm. The variables in the middle three columns are calculated in each of the rest frames of the particles listed in parentheses, and all four calculated values are used as inputs [5].

3.7 DeepAK8

The DeepAK8 classifier arose from an early exploration into the use of DNNs for jet tagging. Like BEST, DeepAK8 can be used to classify heavy particle decays such as those originating from t quarks and $W^{\pm}$, Z, and H bosons against QCD background events. The jets are labeled based on their substructure components including, in particular, any number of b subjets. Furthermore, due the convolutional layers used in the architecture of the DeepAK8 DNN, DeepAK8 is can be classified as a convolutional neural network (CNN) [21, 22].

3.7.1 DeepAK8 Inputs

The input variables for the DeepAK8 neural network are features of the PF jet constituents. PF candidates used include those which can be charged or neutral and particles from reconstructed interactions subsidiary to the main decay (secondary vertices). In total, inputs include no more than 100 inclusive PF candidates, no more than 60 charged PF candidates, and no more than 5 secondary vertices. The inclusive PF candidates are sorted by p_T in descending order while the secondary vertices are ordered by the significance of the impact parameters. Charged PF candidate features (36 overall), inclusive PF candidate features (16 overall), and secondary vertex features (17 overall) used as inputs to DeepAK8 include angular, momentum-based, energy-based, and tracker-based features, and can be seen in full in CMS-AN-17-188 [21].

3.7.2 DeepAK8 Neural Network Architecture

The version of DeepAK8 used in this paper used 3x1 convolutions to deal with the large input size. Separate sets of convolutional layers were used for the inclusive PF candidates (14 convolutional layers), charged PF candidates (14 convolutional layers), and secondary vertices (10 convolutional layers), and were then joined together in one fully connected layer. The activation function, as with BEST is a rectified linear activation function. A schematic of the network can be seen in Figure 3.3. It was trained with the Adam optimizer and implemented with the MXNet package. The samples used were of simulated AK8 jets of p_T greater than 300 GeV and based on CMS Run II data and were made in Summer 2016. For more information on the DeepAK8 neural network, its training, its samples, and its performance please refer to CMS AN-17-188 [21].

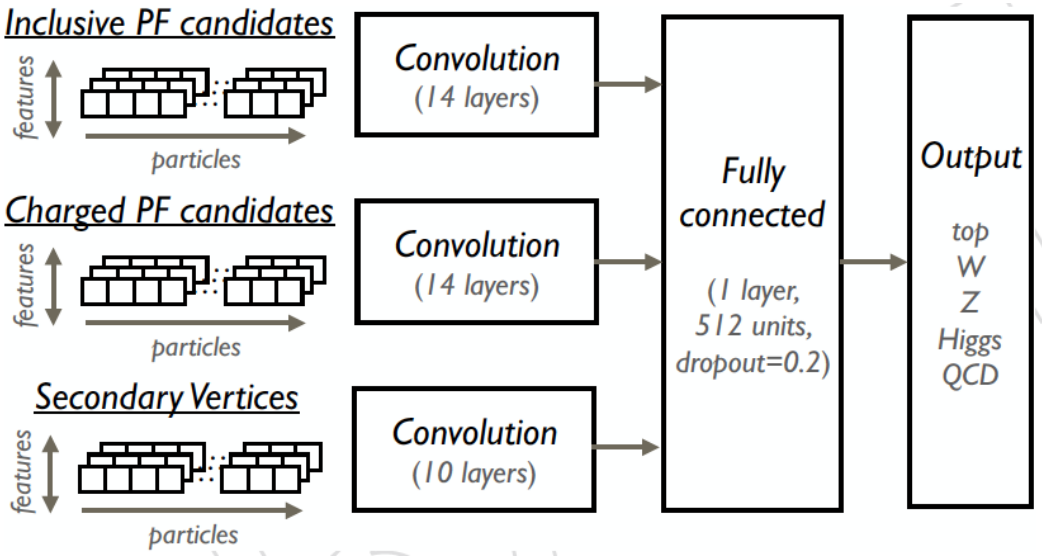


FIGURE 3.3: A diagram showing the architecture of the DeepAK8 neural network [21].

Chapter 4

Data and Sample Processing

4.1 $t\bar{t}$ and QCD Datasets

Data collected by the CMS detector is distributed to users for analysis as MiniAOD (Miniature Advanced Optical Disc) files in the Event Data Model (EDM) format [23]. As this study was concerned with testing a new analysis tool, we used Monte-Carlo (MC) generated $t\bar{t}$ and QCD event samples of the same MiniAOD file type.[1] These datasets were created in Spring 2017 using the POWHEG MC package for calculating jet production in $p\bar{p}$ collisions and interfaced with the PYTHIA8 shower MC generator [24]. They were simulated using GEANT detector simulator package assuming the Run II, Phase II CMS detector upgrade which is set to take data in 2023 during the running of the High-Luminosity LHC (HL-LHC). Both datasets assumed $\sqrt{s}$ of 14 TeV and estimated pileup events, or other physics events with signatures which overlap with those of the particular event being simulated, to be 200 pileup events per event of interest. The $t\bar{t}$ dataset only included events with $m_{t\bar{t}}$ of at least 1500 GeV while the QCD dataset instead made cuts based on p_T and only included events with p_T between 15 and 7000 GeV. Overall, the $t\bar{t}$ dataset consisted of 258 files and included a total of $1,909,854$ events, taking up 490.9 GB of data, while the QCD dataset consisted of 144 files and included a total of $935,590$ events, taking up 334.5 GB of data.

4.2 $t\bar{t}$ and QCD Sample Processing with LJMet

The MiniAOD data or MC sample files are not initially in a form conducive to analysis. Instead, they must be processed into ROOT files containing a flattened "tree" with

[1] The $t\bar{t}$ dataset used was:
\TT_Mtt1500toInf_TuneCUETP8M1_14TeV-powheg-pythia8\PhaseIITDRSpring17MiniAOD-PU200_91X_upgrade2023_realistic_v3-v2\MINIAODSIM
The QCD dataset used was:
\QCD_Flat_Pt-15to7000_TuneCUETP8M1_14TeV_pythia8\PhaseIITDRSpring17MiniAOD-PU200_91X_upgrade2023_realistic_v3-v1\MINIAODSIM

LJMet Jet Cuts	
Variable	**Cut Value**
min jet p_T	15.0 GeV
max jet η	3.0
min AK8 jet p_T	400.0 GeV
max AK8 jet η	2.4
min AK8 jets	2 (QCD), 1 ($t\bar{t}$)
min jets	1
max jets	4000
leading jet p_T	15.0 GeV

TABLE 4.1: Table showing the jet cuts made when processing the initial MiniAOD MC samples into ROOT files. The minimum number of AK8 jets was different for the QCD and $t\bar{t}$ files because we only looked at one leading jet for $t\bar{t}$ samples, but we looked at leading and sub-leading jets in QCD samples to increase statistics.

"branches" corresponding to analysis variables. One tool which can process these files in this way is LJMet, an internal CMS program [23]. LJMet is advantageous due to its versatility and the user's ability to customize the settings and cuts so as to not only calculate and populate specific variables into the final trees, but also make different variable-based event selections during the process. For instance, the main cuts we made while processing our sample datasets into ROOT files pertained to the jets in each event and can be seen in Table 4.1. We also only accepted events in which both top decays were hadronic, thus signifying a fully reconstructable $t\bar{t}$ event.

In addition to the event selection jet cuts made through LJMet, we also used this tool to save the input variables to the BEST neural network, which can be seen in Figure 3.1. Furthermore, we used the calculation capabilities of LJMet to connect PF candidates saved into the MiniAOD file to corresponding jet constituents. In this way, we were effectively able to also save variables similar to some of those used as input features to the DeepAK8 neural network as well. For each jet constituent within each jet, we then saved the constituent's charge, pseudorapidity (η), p_T, energy, and azimuthal scattering angle (ϕ). Finally, we saved the charge, η, p_T, energy, and ϕ of each overall jet.

4.3 From ROOT Trees to DNN Inputs

The final step in preparing the MC samples for input into our DNN was converting the ROOT trees in the ROOT files produced by LJMet into tensors in Hierarchical Data Format (HDF5) files. This process also involved randomly combining and shuffling jets from $t\bar{t}$

events and QCD events. For the combined network of PF and BEST variable inputs, the vectors created (one per jet) were one-dimensional and of size $5 * 11 + 65 = 120$. This structure accounts for the 5 saved pieces of PF feature information (charge, η, p_T, energy, and ϕ) for the overall jet and up to 10 jet constituents. If fewer than 10 jet constituents could be found in a given event, the additional places would be populated with 0s. The additional 65 units of the input tensor account for BEST variables saved for each jet.

By keeping the leading and sub-leading jets from QCD events, we were able to use $449,341$ jets (with approximately equal numbers of $t\bar{t}$ and QCD jets) to train, validate, and test our neural network. Without the additional sub-leading QCD jets, only $277,241$ jets could be used. The final HDF5 input file for the network training on PF features and BEST variables was 168 MB while the input files for training only on PF information or only BEST variables were 85 MB each.

Chapter 5

PF+BEST: A Hybrid Classifier

5.1 Initial Combination Approach

While the final process of combining the BEST and DeepAK8 neural networks is described above, this procedure was the result of multiple attempts to incorporate the two networks. Before settling on the approach documented here, we explored a few other avenues, such as training one of the networks on both BEST and DeepAK8 input variables. The current approach of creating a new DNN which considers BEST and DeepAK8 was chosen due to the limitations of the BEST and DeepAK8 architectures in changing input formats and subsequently retraining the networks. While this approach can also be improved, we believe it is the best way to initially explore the advantages and disadvantages of integrating features of both the BEST and DeepAK8 taggers.

5.2 PF+BEST Network Architecture

The final combination of DeepAK8 and BEST, referred to as PF+BEST due to its input variables including PF information and BEST kinematic and topological variables, was in the form of a convolutional neural network with two, one-dimensional convolutional layers, four "main" dense layers, and one final dense layer. The activation function used was a rectified linear activation function until the final step, when a sigmoid activation function was used to set the outputs between 0 and 1.

The PF+BEST network was trained on a total of $268,800$ jets, validated on a set of $89,600$ jets, and tested on a set of $89,600$ jets for a 3:1:1 training-validation-testing ratio. There was no overlap between the training, testing, and validation sets. Further information on the structure of the PF+BEST neural network can be seen in Figure 5.1.

PF+BEST was created, trained, and tested with the Keras Python library with a TensorFlow Python package backend [25, 26]. The scikit-learn Python package was used to create Receiver Operating Characteristic (ROC) curves and calculate their Area Under the

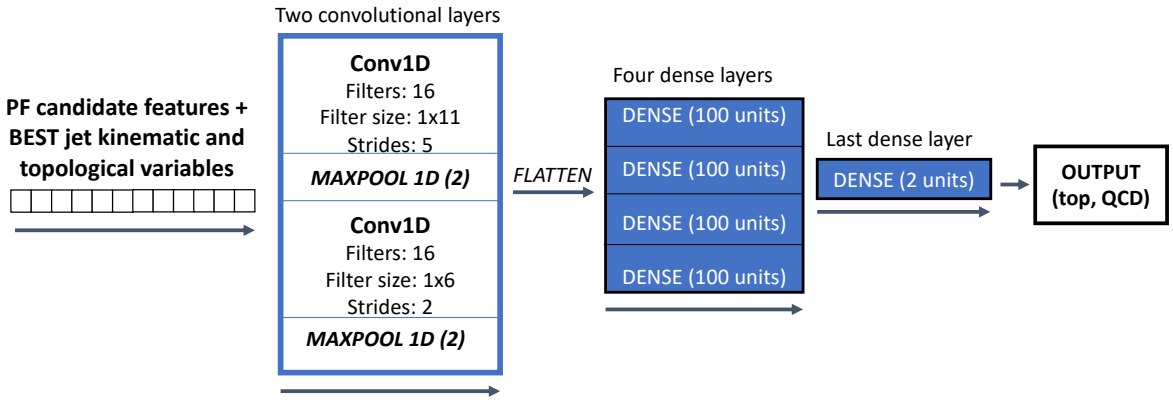


FIGURE 5.1: A schematic showing the general architecture of the PF+BEST neural network. This architecture was modified in a few tests, but this base network, with two convolutional layers, four "main" dense layers, and one final dense layer, was the primary structure used for training and testing.

Curve (AUC) scores. ROC curves and AUC scores can be used to understand the relationship between the true and false positive rates of the classifier. Network training, testing, and production of ROC curves was done remotely on graphics processing units (GPUs) located at the Fermi National Accelerator Laboratory in Batavia, Illinois. PF+BEST networks and variations were trained for between 25 and 75 epochs. As each epoch took between 60 and 120 seconds to train, network training time ranged from approximately 25 minutes to approximately 2.5 hours. The time taken to evaluate the trained networks by running the testing datasets through them was approximately less than 2 minutes. Finally, in order to avoid overtraining, the epoch used for testing was chosen based on the best (largest) AUC score achieved before the AUC scores significantly leveled out, as determined by the author.

Chapter 6

Results

6.1 Initial PF+BEST and Increasing Statistics

The initial version of of PF+BEST was run on 277,241 instead of the 449,341 jets used in final results. The differences in the ROC curves and AUC scores between these two sets can be seen in Figure 6.1. It is important to note that these initial results trained and tested on files with extra BEST-related input variables giving the number of jets in the frames of the detector and t, W, Z, and H at rest, as well as jet mass and energy, for a total BEST variable count of 73 instead of 65. These variables seem to have been left over in the BEST files from LJMet for unknown reasons, but should not necessarily be part of the variables on which a BEST-type network would be trained. While they were removed from training, testing, and validation sets for subsequent versions of PF+BEST, our increase in sample statistics occurred before this change and thus is represented in this initial form.

6.2 Input-Varying Versions of PF+BEST

While PF+BEST nominally includes PF candidate feature information and BEST kinematic and topological jet variable information, we also trained identical networks only using our PF candidate feature information (PF_CNN) or only our BEST kinematic and topological jet variables (BEST_CNN). The ROC curves comparing these results with PF+BEST can be seen in Figure 6.2.

6.3 Comparing PF+BEST, BEST, and DeepAK8

In addition to comparing the architecture of PF+BEST as trained on different subsets of inputs, we can also compare the performance of PF+BEST to the BEST and DeepAK8 classifiers themselves. We also have included the BEST_CNN network from Figures 6.2, which describes the output of the PF+BEST CNN network architecture with only BEST variable inputs, in order to compare the original BEST to this BEST-like network. This can

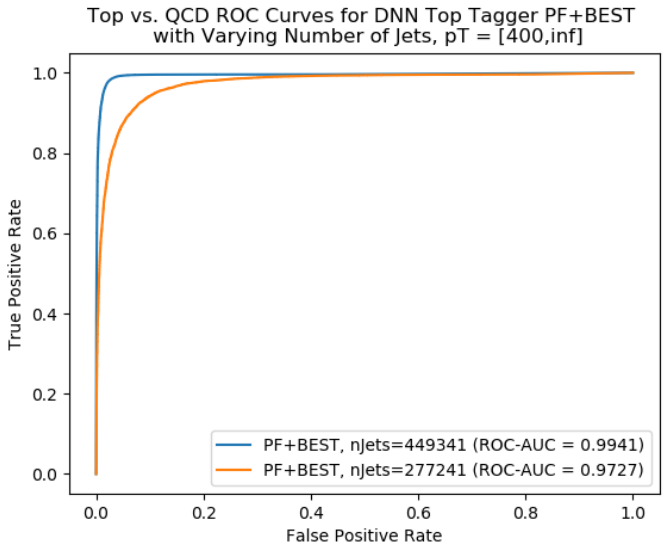
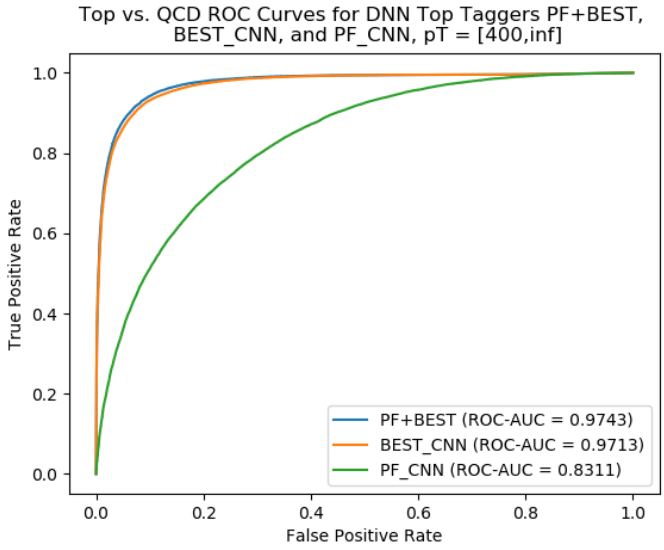


FIGURE 6.1: A comparison of ROC curves corresponding to versions of PF+BEST trained, tested, and validated on different numbers of jets. The PF+BEST network using a sample of 277,241 jets was trained for 40 epochs and evaluated at epoch 31 while the network using a sample of 449,341 jets was trained for 75 epochs and evaluated at epoch 57. The p_T range included is in units of GeV.

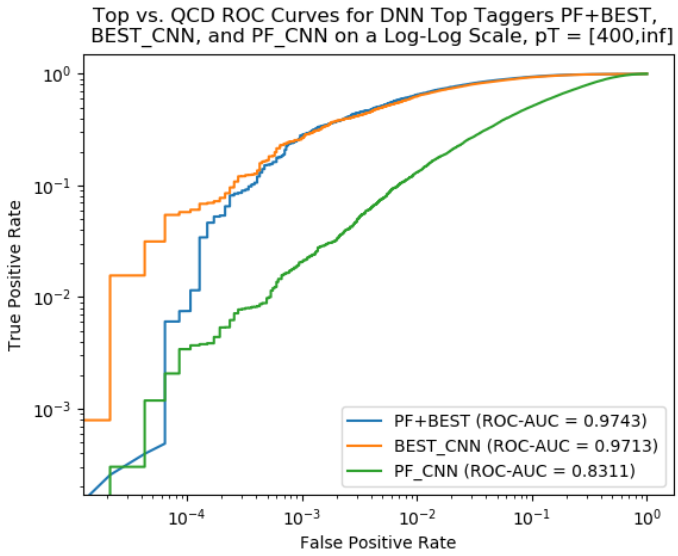
be seen in Figure 6.3 for the entire p_T range of 400 GeV to infinity and in five p_T ranges of 200 GeV each in Figures 6.4 (linear plots) and 6.5 (log-log plots).

6.4 Testing Different Architectures for PF+BEST

While PF+BEST was primarily used with the architecture described in Chapter 5, it was also tested with two other architectures. First, the number of layers was cut in half, so only the first convolutional layer and first two "main" dense layers were used. Additionally, the number of layers was doubled such that four convolutional layers and eight "main" dense layers were used. The four additional dense layers were identical to the original dense layers, while the two additional convolutional layers had the same number of filters, but the filters were sized 1x3 (first additional layer) and 1x2 (second additional layer) and each had strides of 1. The performance of these networks can be seen in Figure 6.6.

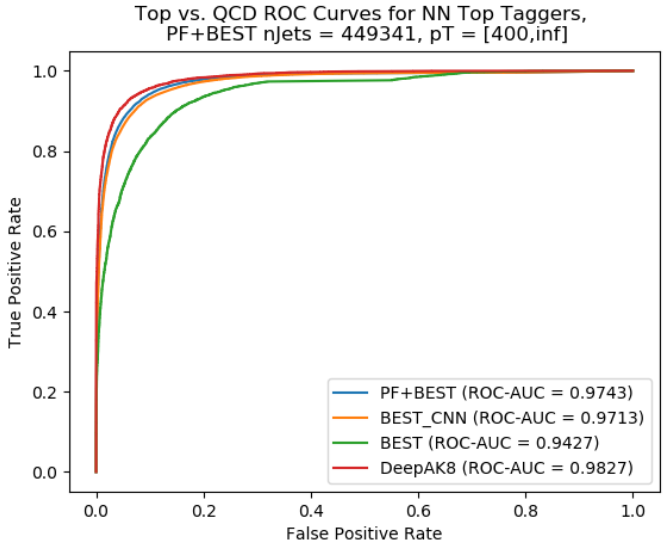


(A)

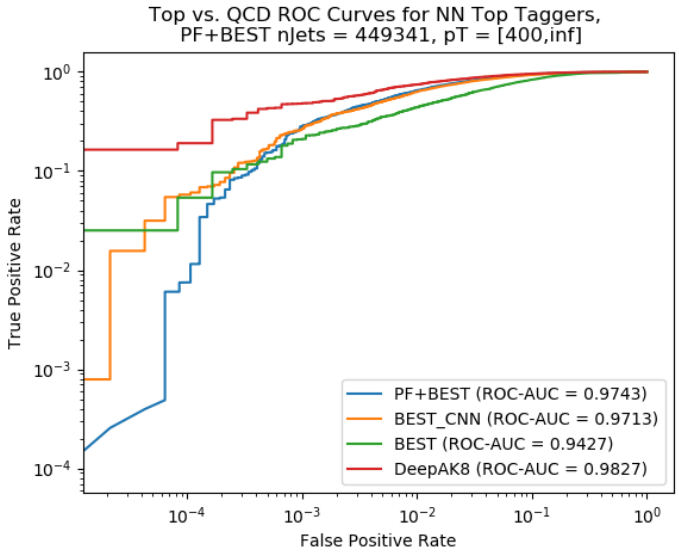


(B)

FIGURE 6.2: Comparisons of ROC curves corresponding to versions of PF+BEST, which was trained on PF information and BEST variables, PF_CNN, which was trained on only PF information, and BEST_CNN, which was only trained on BEST variables, on both linear (6.2a) and log-log scales (6.2b). Apart from the dimensions of the input tensors (1x120, 1x55, and 1x65 respectively), all network architectures were identical. The PF+BEST and BEST_CNN networks were each trained for 75 epochs and evaluated at epoch 30 while the PF_CNN network was trained for 50 epochs and evaluated at epoch 46. The p_T range included is in units of GeV.



(A)



(B)

FIGURE 6.3: Comparisons of ROC curves corresponding to PF+BEST, BEST_CNN, BEST, and DeepAK8 on both linear (6.3a) and log-log scales (6.3b). PF+BEST and BEST_CNN were each trained for 75 epochs and evaluated at epoch 30 while BEST and DeepAK8 were trained independent of this project by other authors as described in Chapter 3. The p_T range included is in units of GeV.

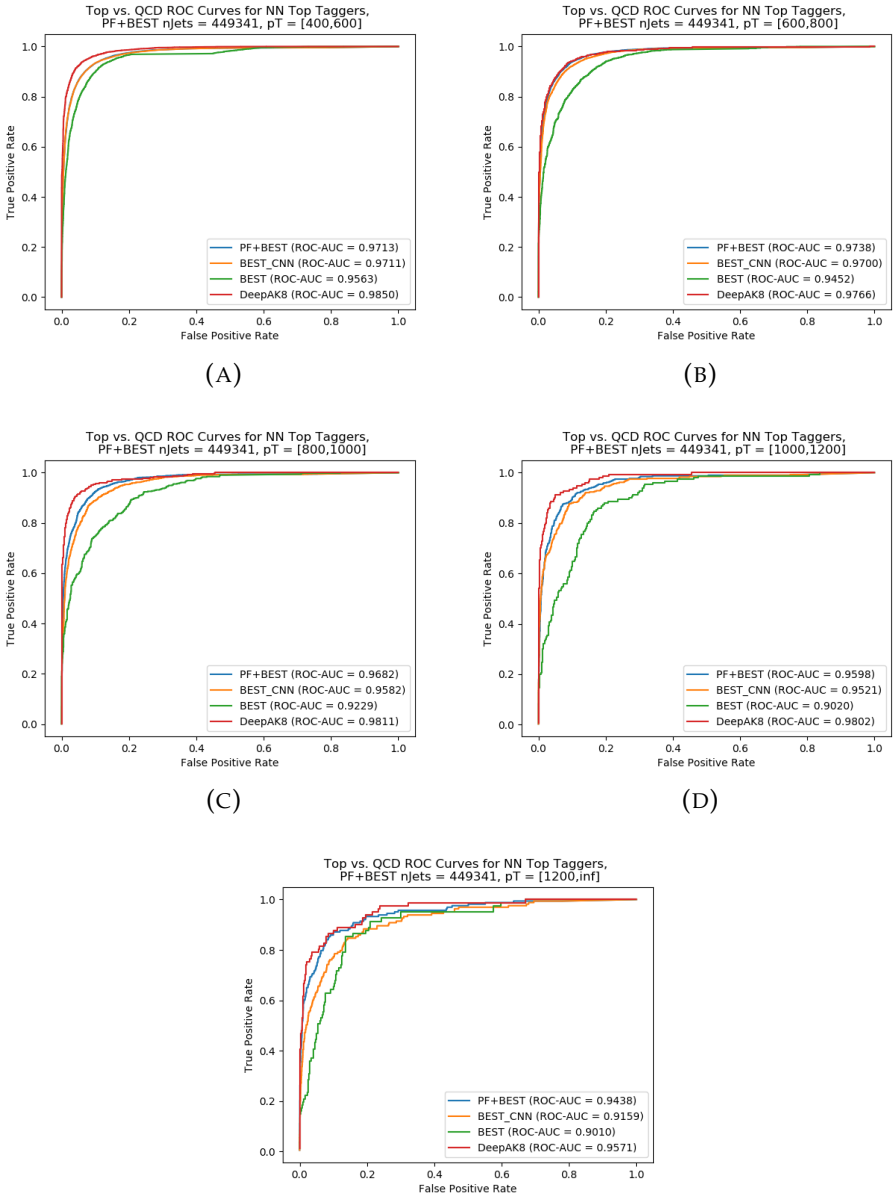


FIGURE 6.4: Comparisons of ROC curves corresponding to PF+BEST, BEST_CNN, BEST, and DeepAK8 on linear scales in five, 200 GeV-sized p_T ranges between 400 GeV to infinity. PF+BEST and BEST_CNN were each trained for 75 epochs and evaluated at epoch 30 while BEST and DeepAK8 were trained independent of this project by other authors as described in Chapter 3. The p_T range included is in units of GeV.

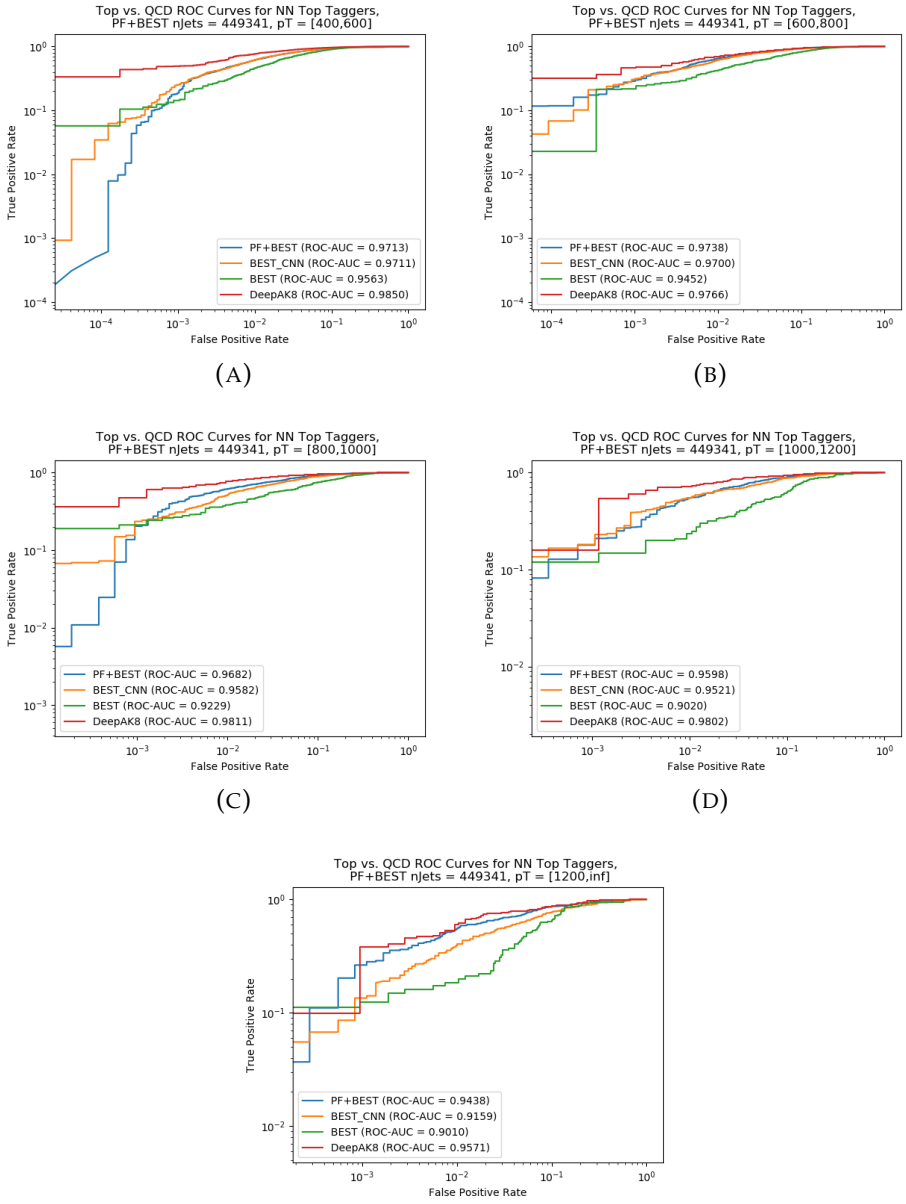
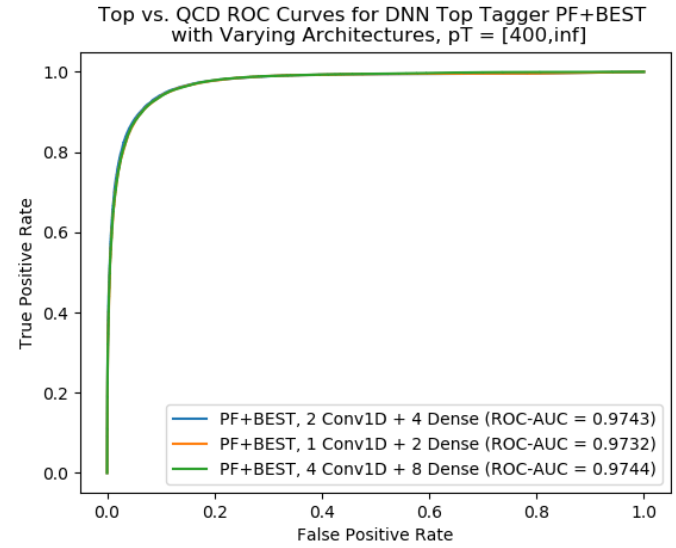
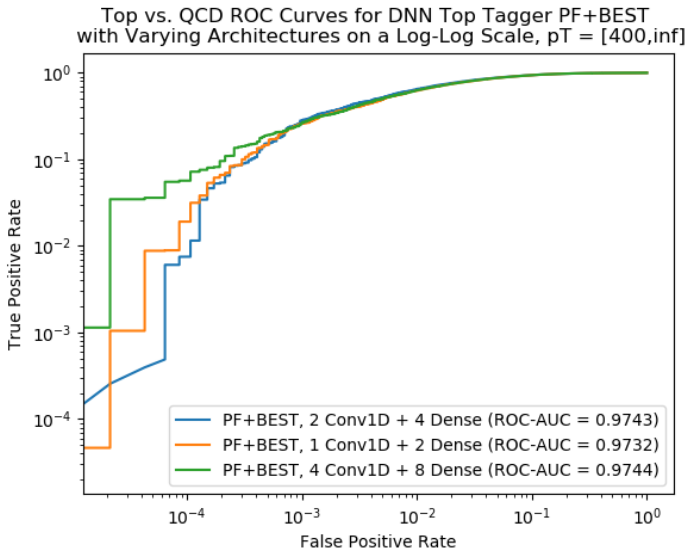


FIGURE 6.5: Comparisons of ROC curves corresponding to PF+BEST, BEST_CNN, BEST, and DeepAK8 on log-log scales in five, 200 GeV-sized p_T ranges between 400 GeV to infinity. PF+BEST and BEST_CNN were trained for 75 epochs each while BEST and DeepAK8 were trained independent of this project by other authors as described in Chapter 3. The p_T range included is in units of GeV.



(A)



(B)

FIGURE 6.6: Comparisons of ROC curves corresponding to versions of PF+BEST of varying neural network architectures on both linear (6.6a) and log-log scales (6.6b). The original PF+BEST network (2 Conv1D + 4 Dense) was trained for 75 epochs and evaluated at epoch 30 while the two other architectures (1 Conv1D + 2 Dense and 4 Conv1D + 8 Dense) were both trained for 25 epochs and evaluated at epoch 22. The p_T range included is in units of GeV.

Chapter 7

Discussion and Analysis

7.1 Evaluating PF+BEST

Overall, our inquiry into combining aspects of BEST and DeepAK8 to create a more effective DNN boosted top tagger gave encouraging results. We can see this by observing the ROC curves and AUC scores of the different networks tested with the knowledge that an AUC score of 1.0, representing a true positive rate of 1.0 for a false positive rate of 0.0, is the ideal AUC score. In this way, we can evaluate PF+BEST through variations in networks based on statistics, inputs, and architectures, as well as in relation to BEST and DeepAK8 over a full p_T spectrum and over five p_T bins. As this was a preliminary study, we did not train each network multiple times and thus do not know definitively the difference in AUC scores which would be statistically significant. This is definitely an area which can be improved in future studies but, for this study, we note all differences in AUC score with the knowledge that those under 1% (0.01) and especially those under 0.5% (0.005) may not be statistically significant.

7.2 Increasing Statistics

As seen in Figure 6.1, we were able to effectively increase our sample of jets by including leading and sub-leading QCD jets in order to increase our statistics and the strength of PF+BEST. By increasing the number of jets from $277,241$ to $449,341$, we were able to increase the AUC score by over 0.02, though it is important to recognize that this particular improvement was achieved when still using samples with the extraneous BEST-related input variables described in Chapter 6.1. Furthermore, increasing sample statistics is also helpful in preventing any overtraining of the neural network.

7.3 Varying Inputs

In addition to successfully increasing statistics and identifying extraneous BEST-related input variables, we were able to see that PF+BEST was more effective than identically structured networks trained only on BEST input variables (BEST_CNN) or only on PF level information we selected (PF_CNN), as seen in Figure 6.2. From Figure 6.2a, we can clearly see that the AUC scores of PF+BEST and BEST_CNN, 0.9743 and 0.9713 respectively, easily surpass the PF_CNN AUC score of 0.8311.

The weakness of PF information only is expected due to the fact that PF+BEST and, thus, PF_CNN, only trained on five PF candidate features. The efficacy of DeepAK8 on which this project was predicated were the result of many more candidate features. It is interesting, however, that BEST_CNN performs so similarly to PF+BEST, with a difference in AUC score of only 0.003, which could only be due to statistical fluctuations. The difference between the two networks can be seen more closely in Figure 6.2b, in which it is clear that BEST_CNN has a higher true positive rate at lower false positive rates, which is desirable, but that PF+BEST steeply increases such that at higher true positive rates, its false positive rate is slightly lower than that of BEST_CNN, leading to its overall higher AUC score. The differences between BEST_CNN and PF+BEST are further discussed in the following section, Chapter 7.4.

From these observations of Figure 6.2 as a whole, it is clear that training on so few PF features does not significantly help the accuracy of PF+BEST versus BEST_CNN. The slightly higher AUC score of PF+BEST, however, might suggest that training PF+BEST on a greater number of PF features in future studies could increase the difference in performance of PF+BEST versus BEST_CNN.

7.4 Direct Comparisons of PF+BEST, BEST_CNN, BEST, and DeepAK8 in Full p_T Spectrum

PF+BEST also performed comparably to BEST and DeepAK8 in direct comparisons to these two networks. In particular, we can see from Figure 6.3 that both PF+BEST and BEST_CNN performed better than BEST and only slightly worse than DeepAK8, with AUC scores of 0.9743, 0.9713, 0.9427, and 0.9827 respectively, over the full p_T spectrum of 400 GeV to infinity.

While PF+BEST, DeepAK8, and BEST_CNN all seem relatively comparable in Figure 6.3a, however, it is interesting to look at the log-log plot of Figure 6.3b, which shows that PF+BEST once again has the lowest true positive rate at very low false positive rates, but goes on to increase dramatically such that its overall AUC score is the second highest of the four represented. Furthermore, it is BEST and DeepAK8 which have the highest two true positive rates at the lowest false positive rates. It will be important in future studies

and testing of PF+BEST to investigate why these differences, only apparent in the log-log ROC curve graphs, occur between PF+BEST and BEST_CNN, and more established networks such as DeepAK8 and BEST.

Furthermore, one caveat that should be recognized in the direct comparisons of these networks is that both BEST and DeepAK8 were trained on slightly different samples, both with less pileup than the PF+BEST samples [5, 17, 21, 22]. The networks were still all tested in the PF+BEST samples referenced in Chapter 4.1, but the fact that they were trained on different sets may explain, for instance, why BEST performed worse than the other two networks. Alternatively, BEST's worse performance as compared to PF+BEST and BEST_CNN may be due to the simpler, non-convolutional architecture of BEST as compared to these two networks [5, 17].

7.5 Direct Comparisons of PF+BEST, BEST_CNN, BEST, and DeepAK8 in Varying p_T Bins

In addition to comparing PF+BEST, BEST_CNN, BEST, and DeepAK8 over the entire p_T spectrum studied, it is important to see if there are any differences in the way in which the networks compare in smaller p_T ranges, as seen in Figures 6.4 and 6.5. Overall, the performance of PF+BEST was best in the 600 to 800 GeV range, as seen most clearly in Figure 6.5b. In this p_T range, the AUC score for PF+BEST was only 0.0028 lower than that of DeepAK8, and its initial true positive rate was second highest only to DeepAK8.

The other interesting result in these plots was the superiority of PF+BEST in relation to BEST_CNN in high p_T ranges. This is clear from the decrease in BEST_CNN's AUC score from 0.9521 in the 1000 to 1200 GeV p_T range to 0.9159 in the highest p_T range of 1200 GeV to infinity, while PF+BEST's AUC scores only decreases from 0.9598 to 0.9438 between these two p_T ranges. Prior to that last bin, the performance of BEST_CNN was very comparable to PF+BEST, which was still generally better than BEST, though in the last p_T bin BEST_CNN seemed to act much more similarly to BEST than PF+BEST. The lower AUC scores for BEST on its own at higher p_T values may be related to the fact that BEST was trained on samples which only went up to p_Ts of 2 TeV while the other networks were not trained on samples with upper p_T limits [5, 21]. The fact, however, that the performance of BEST declines before the highest p_T bin and that BEST_CNN's performance also seems to decline at higher p_T values may be related to a lower efficacy in using solely those variables to tag jets at higher p_Ts as opposed to using other or additional inputs like DeepAK8 or PF+BEST. The superiority of DeepAK8 and PF+BEST in comparison to BEST and BEST_CNN is therefore important to probe in further studies as many forms of BSM physics might result in very high p_T jets.

7.6 PF+BEST with Varying Architectures

While we did not test a large variety of architectures for PF+BEST, we did test effectively doubling and halving the number of hidden layers in PF+BEST, as seen in Figure 6.6. From Figure 6.6a, it is virtually impossible to tell the differences between the three architectures. Their AUC scores fall within 0.0012 of each other, with the "original" PF+BEST having an AUC score of 0.9743, PF+BEST with additional layers having an AUC score of 0.9744, and PF+BEST with fewer layers having an AUC score of 0.9732. From Figure 6.6b, we can see the slightly higher true positive rate of PF+BEST with additional layers at the lowest false positive rate, but the differences between the three networks are not truly significant. It would, however, be interesting to look at these networks in different p_T ranges to see if performance changes based on p_T bin. Furthermore, our variances in architecture were very preliminary and could be significantly diversified by trying even more layers, different proportions of convolutional and dense layers, different sizes of dense layers, different filter numbers, filter size, and stride lengths for the convolutional layers, and many additional variations. In this way, the testing of different architectures is a particularly rich avenue of exploration for future studies of PF+BEST-like DNNs.

Chapter 8

Conclusions and Moving Forward

8.1 Does PF+BEST have a future?

The initial goal of creating PF+BEST was to investigate DNN boosted top tagging techniques which might act as alternatives to more established networks such as BEST and DeepAK8. Overall, it was found that PF+BEST is comparable with these two established networks and has the possibility, with further studies, to improve upon its performance here and either become equal with or even surpass the performance of the two networks. It should also be noted that PF+BEST is currently relatively simple, especially compared to a network such as DeepAK8, which is important when considering the time and computing resources necessary to run such a network on a large scale.

Furthermore, the fact that initial tests of PF+BEST was able to achieve these results for files with pileup values of 200 is extremely significant. These pileup values are expected at the HL-LHC and are 4 to 5 times higher than those seen during Run II of the LHC between the years of 2016 and 2019. Unlike PF+BEST, the BEST and DeepAK8 networks were trained on MC samples produced for Run II analyses and, therefore, were not trained to be used on data MC samples with 200 pileup [5, 21]. Future top taggers must be able to combat these challenging, high pileup regimes in order to identify top quarks in the HL-LHC, making the efficiency and efficacy of PF+BEST exciting for the future of top tagging. For these reasons, it is clear that the hadronic top tagging capacities of PF+BEST should be further explored in future studies.

8.2 Next Steps for PF+BEST

It is first important to consider future developments of PF+BEST along avenues described in Chapter 7. For instance, further increases in sample statistics, testing different DNN architectures and adding more PF features as inputs could lead to significant improvements of the network. Furthermore, explorations into differences in performance between BEST,

PF+BEST, and DeepAK8 in different p_T ranges could give further insight into improvements on these networks. It would also be helpful in future studies to collaborate more closely with the creators of BEST and DeepAK8 such that BEST, DeepAK8, and PF+BEST can be all trained and tested on the same samples. Finally, while we have clearly established the importance of top tagging, it would be interesting to try to extend PF+BEST such that it, like current versions of BEST and DeepAK8, is able to tag other heavy particles, such as $W^{\pm}$, Z, and H bosons, in addition to top quarks. If PF+BEST is developed along these guidelines, there is significant promise that PF+BEST can play an important role in discovering new physics involving coupling with heavy SM particles in the future.

Bibliography

[1] Mark Thomson. *Modern Particle Physics*. Cambridge: Cambridge University Press, 2016.

[2] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020.

[3] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021.

[4] *The Standard Model and beyond*. URL: http://united-states.cern/physics/standard-model-and-beyond (visited on 04/15/2019).

[5] Justin Pilot et al. "Search for Pair Production of Vector-Like T Quarks in the Fully Hadronic Channel with the Boosted Event Shape Tagging Algorithm". In: (2018). CMS AN-18-095.

[6] Martin Schmaltz and David Tucker-Smith. "Little higgs theories". In: *Annu. Rev. Nucl. Part. Sci.* 55 (2005), pp. 229–270. DOI: 10.1146/annurev.nucl.55.090704.151502.

[7] Lily Asquith et al. *Jet substructure at the Large Hadron Collider: experimental review*. 2018. URL: https://arxiv.org/abs/1803.06991.

[8] Albert M Sirunyan et al. "Measurement of vector boson scattering and constraints on anomalous quartic couplings from events with four leptons and two jets in proton–proton collisions at s= 13TeV". In: *Physics letters B* 774 (2017), pp. 682–705. DOI: 10.1016/j.physletb.2017.10.020.

[9] *CMS Detector*. URL: https://cms.cern/detector (visited on 04/15/2019).

[10] Albert M Sirunyan, CMS Collaboration, et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/p10003.

[11] Tilman Plehn and Michael Spannowsky. "Top tagging". In: *Journal of Physics G: Nuclear and Particle Physics* 39.8 (2012), p. 083001. DOI: 10.1088/0954-3899/39/8/083001.

[12] Carmen Diez Pardos. *Top Quark Physics at ATLAS and CMS*. PDF. Presented at Corfu2017: Workshop on the Standard Model and Beyond in Corfu, Greece, September 2-10, 2017. 2017. URL: http://www.physics.ntua.gr/corfu2017/Talks/carmen_diez@desy_de_01.pdf.

[13] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (2008), p. 063. URL: https://arxiv.org/pdf/0802.1189.pdf.

[14] Jyothsna Rani Komaragiri for the CMS collaboration. *Physics objects for top physics in CMS*. Presented at the 9th International Workshop on Top Quark Physics in Olomouc, Czech Republic, September 19–23, 2016. 2016. URL: https://arxiv.org/abs/1612.01917.

[15] Mrinal Dasgupta et al. "Towards an understanding of jet substructure". In: *Journal of High Energy Physics* 2013.9 (2013), p. 29. DOI: 10.1007/JHEP09(2013)029.

[16] Andrew J Larkoski, Ian Moult, and Benjamin Nachman. "Jet substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning". In: *arXiv preprint arXiv:1709.04464* (2017). URL: https://arxiv.org/pdf/1709.04464.pdf.

[17] JS Conway et al. "Identification of high-momentum top quarks, Higgs bosons, and W and Z bosons using boosted event shapes". In: *Physical Review D* 94.9 (2016), p. 094027. DOI: 10.1103/PhysRevD.94.094027.

[18] Jesse Thaler and Ken Van Tilburg. "Identifying boosted objects with N-subjettiness". In: *Journal of High Energy Physics* 2011.3 (2011), p. 15. DOI: 10.1007/JHEP03(2011)015.

[19] Albert M Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* (2018). DOI: 10.1088/1748-0221/13/05/p05011.

[20] Andrew J Larkoski et al. "Soft drop". In: *Journal of High Energy Physics* 2014.5 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146.

[21] The CMS Collaboration. "Deep learning for jet reconstruction". In: (2017). CMS AN-17-188.

[22] Markus Stoye et al. "DeepJet: Generic physics object based jet multiclass classification for LHC experiments". In: *Proceedings of the Deep Learning for Physical Sciences Workshop at NIPS*. 2017. URL: https://dl4physicalsciences.github.io/files/nips_dlps_2017_10.pdf.

[23] Julie Hogan. *LJMet Tutorial*. unpublished. Presented as tutorial to Brown University CMS Group students in October 2018. 2018.

[24] J.M. Campbell et al. "Top-pair production and decay at NLO matched with parton showers". In: (2015). DOI: 10.1007/JHEP04(2015)114.

[25] François Chollet et al. *Keras*. https://keras.io. 2015.

[26] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.