



Electron Identification with a Convolutional Neural Network in the ATLAS Experiment

The ATLAS Collaboration

The identification of electrons plays an important role for a large fraction of the physics analyses performed at the ATLAS experiment. An improved electron identification algorithm is presented that is based on convolutional neural networks (CNN), a type of machine learning architecture specialized in image recognition. It takes as input the images of the deposited energy in the calorimeter cells around the reconstructed electron candidates for each of the electromagnetic and hadronic calorimeter layers. Additional input features include the same high-level variables that are used by the likelihood (LLH) and deep neural network (DNN) algorithms developed in ATLAS, as well as the information of up to five inner detector tracks that are matched to an electron candidate during its reconstruction. The output of the network corresponds to the probability that a reconstructed electron belongs to six classes of signals and backgrounds. A significant improvement in identification performance is observed when the CNN algorithm is used in the simulation. For example, for a working point that corresponds to the same signal efficiency as the LLH "Loose" working point, the CNN improves the rejection against charged hadrons faking signal electrons, the dominant electron background at the LHC, by factors of 5 to 8 (depending on the electron kinematics) with respect to the LLH. For the most difficult background constituted of electrons originating from heavy flavour hadron decays, the background rejection of the CNN is improved by factors varying between about 2 to 3.5 with respect to the LLH.

Contents

1	Introduction	2
2	ATLAS Detector	3
3	Simulated samples and electron candidate classification	5
3.1	Classification of electron signals and backgrounds	5
3.2	Simulated samples used for CNN training and performance evaluation	6
4	Input variables to the CNN	7
4.1	Electron reconstruction	8
4.2	High-level variables	8
4.3	Additional track variables	8
4.4	Calorimeter images	8
5	Description of the CNN	11
5.1	Preparation of the Training Sample	11
5.2	Preprocessing	11
5.3	Network Architecture	12
5.4	Definition of the Final Discriminant	15
6	Performance	16
6.1	Analysis of the performance and comparison with the Likelihood algorithm	16
6.2	Determination of the population of electron classes in a sample using the CNN	20
7	Conclusion	28

1 Introduction

The capability to reliably identify signal electrons represents a pillar of the ATLAS physics program. Electron objects are used in a large fraction of ATLAS publications, whether to select or veto electron candidates. Better identification implemented in the electron triggers would allow to lower their transverse momentum threshold and increase their acceptance. In particular, analyses featuring multiple leptons, pairs of electrons with the same electric charge (known as same-sign lepton pairs) or those needing to identify loosely signal electrons can suffer from large, sometimes dominant, fake or non-prompt electron background. In summary, any improvement to the identification of electrons is expected to benefit a large fraction of the ATLAS analyses planned in the future.

The identification of electrons has been mainly performed so far in ATLAS by the likelihood (LLH) algorithm [1, 2]. This algorithm employs high-level features defined using information from the inner detector (including the particle identification capability of the Transition Radiation Tracker (TRT)), and the electromagnetic and hadronic calorimeters, as well as variables that combine the information from more than one of these subdetectors. These variables are selected to provide good discrimination between signal electrons and electron backgrounds coming from a real data sample that is expected to be dominated by light flavour hadrons faking the signature of signal electrons. Probability density functions for each

of these variables are created for electron signal and background and are combined to obtain a binary likelihood discriminant. The correlations between the high-level features are neglected when forming the likelihood discriminant.

A novel electron identification algorithm has recently been introduced in ATLAS that employs a deep neural network (DNN) [3] that uses as input (mostly) the same high-level features as the LLH but treats them more optimally, in particular the correlations between them, to increase significantly the electron identification performance. Depending on the kinematics of the electron candidate, an increase in combined background rejection between 1.7 and 5.5 is obtained in simulated data for a fixed signal efficiency. In addition to the identification performance, the DNN algorithm provides multinomial classification, which means that the electron backgrounds are separated in several classes such as charged hadrons faking electrons or electrons from heavy-flavor decays. Such classification provides the flexibility to optimize the electron background rejection for analyses that particularly suffer from one of these background classes.

More optimal electron identification can be achieved by using the low level information that the ATLAS detector provides, such as the energy deposited in the individual cells of each of the calorimeter layers around an electron candidate, which can be represented as images [4]. In the following, these calorimeter images are treated by a convolutional neural network (CNN) [5, 6], a type of neural network architecture that is specialized in image recognition. To obtain optimal performance, the information of the calorimeter images is combined with high-level features similar to those used in the LLH and DNN algorithms. This CNN algorithm also uses additional tracks that are matched to electron candidates, which are more numerous on average for background than for signal electrons.

The design and performance of the electron identification CNN are described in this note, which will focus on electrons with $p_T > 4.5$ GeV (where p_T is the electron momentum transverse to the LHC beam) and $|\eta| < 2.5$ ¹. The ATLAS detector is presented in Section 2. The electron candidate classification and the samples used for training the CNN are described in Section 3. Section 4 details the inputs to the CNN. Section 5 describes the neural network architectures and the training procedure. Finally, the performance of the CNN is compared with the existing algorithms using simulated samples in Section 6 before concluding remarks are provided in Section 7.

2 ATLAS Detector

The ATLAS detector [7] at the LHC covers nearly the entire solid angle around the collision point. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets.

The inner-detector system (ID) is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit normally being in the insertable B-layer (IBL) installed before Run 2 [8, 9]. It is followed by the silicon microstrip tracker (SCT), which usually provides

¹ ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z -axis along the beam pipe. The x -axis points from the IP to the centre of the LHC ring, and the y -axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z -axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The TRT also provides electron identification information based on the fraction of hits (typically 30 in total) above a higher energy-deposit threshold corresponding to transition radiation.

The ATLAS calorimeter system has both electromagnetic (EM) and hadronic (HAD) components and covers the pseudorapidity range $|\eta| < 4.9$, with finer granularity over the region matching the inner detector. The central electromagnetic calorimeters are of an accordion-geometry design made from lead/liquid-argon (LAr) detectors, providing a full ϕ coverage. These detectors are divided into two half-barrels ($-1.475 < \eta < 0$ and $0 < \eta < 1.475$) and two endcap components ($1.375 < |\eta| < 3.2$), with a transition region between the barrel and the endcap ($1.37 < |\eta| < 1.52$) which contains a relatively large amount of inactive material. Over the region devoted to precision measurements ($|\eta| < 2.5$, excluding the transition regions), the EM calorimeter is segmented into longitudinal compartments, referred to as the first (also known as strips), second, and third layers. The first layer consists of strips finely segmented in $|\eta|$, offering excellent discrimination between photons and $\pi^0 \rightarrow \gamma\gamma$ decays. The granularity of that layer is 0.1 in ϕ , and depends on $|\eta|$ for its granularity in the η direction: 0.025/8 for $|\eta| < 1.8$, 0.025/6 for $1.8 < |\eta| < 2.0$, 0.025/4 for $2.0 < |\eta| < 2.4$ and 0.025 for $2.4 < |\eta| < 2.5$. Most of the energy is typically collected in the second layer, which has a lateral granularity of 0.025×0.025 in (η, ϕ) space, while the third layer provides measurements of energy deposited in the tails of the shower. The granularity of the third layer is 0.050×0.025 in (η, ϕ) . The central EM calorimeter is complemented by two presampler detectors in the region $|\eta| < 1.52$ (barrel) and $1.5 < |\eta| < 1.8$ (endcaps), made of a thin LAr layer, providing a sampling for particles that start showering in front of the EM calorimeters. The granularity of the presampler is 0.025×0.1 in (η, ϕ) .

The hadronic calorimeter, surrounding the EM calorimeter, consists of an iron/scintillator tile calorimeter in the range $|\eta| < 1.7$ and two copper/LAr calorimeters spanning $1.5 < |\eta| < 3.2$, which are each composed of three and four longitudinal layers, respectively. The granularity of the tile calorimeter is 0.1×0.1 in (η, ϕ) , except for its last layer for which it is 0.2×0.1 . The granularity in the LAr HAD endcap calorimeter is 0.1×0.1 in (η, ϕ) for $|\eta| < 2.5$, the range of interest for this note. The acceptance is extended by two copper/LAr and tungsten/LAr forward calorimeters extending up to $|\eta| = 4.9$, and hosted in the same cryostats as the electromagnetic calorimeter.

The muon spectrometer surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. The muon spectrometer includes a system of precision tracking chambers and fast detectors for triggering.

Interesting events are selected by the first-level trigger system implemented in custom hardware, followed by selections made by algorithms implemented in software in the high-level trigger [10]. The first-level trigger accepts events from the 40 MHz bunch crossings at a rate below 100 kHz, which the high-level trigger further reduces in order to record events to disk at about 1 kHz.

An extensive software suite [11] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Simulated samples and electron candidate classification

3.1 Classification of electron signals and backgrounds

The CNN, which is described in detail in Section 5, undergoes supervised training using simulated data where signal and background electrons are assigned a label based on their true identity, which is available in the simulation. This multinomial classification has been designed to provide flexibility to the users, such that analyses suffering from different types of electron backgrounds can modify the computation of the final CNN discriminant to provide optimal background rejection for their specific analysis. This aspect of the CNN is described in more detail in Section 6. The same electrons classes as for the DNN are used.

- **Prompt electrons:** Prompt isolated electrons, e.g. electrons from $Z \rightarrow e^+e^-$, $W^\pm \rightarrow e^\pm \nu$, $J/\psi \rightarrow e^+e^-$ decays with the J/ψ being produced in the hard scatter. Electrons from a final state radiation or bremsstrahlung photons are also considered here if they originate from a prompt electron. Furthermore, the reconstructed charge is the same as the true charge;
- **Charge-flip electrons:** Same as prompt electrons, but the reconstructed charge is the opposite of the true charge. Charge-flip electrons originate either from material effects (e.g. $e^- \rightarrow e^- \gamma \rightarrow e^- e^+ e^-$ for which the positron is reconstructed as the main electron candidate track) or from a pure mistake in the electron curvature measurement, which occurs in particular at high pseudorapidity where the lever arm of the inner detector is reduced, or at high p_T . In case of an electron originating from bremsstrahlung followed by a conversion, the charge of the original prompt electron is considered the true charge. For most use-cases in the ATLAS physics program, charge-flip electrons are considered signals. Notable exceptions are analyses featuring pairs of same-sign leptons. With the DNN or the CNN, the users can decide whether to include charge-flip electrons as signal or background (see Section 6);
- **Prompt photon conversion:** Electrons from prompt photons which convert into an e^+e^- pair. Prompt photons which are reconstructed as an electron are also considered for this class;
- **Electrons from heavy-flavour hadron decays:** Electrons coming from a decay of a c -hadron or a b -hadron, except those produced from prompt J/ψ decays as explained above. Electron backgrounds from this and the next two classes originate from hadrons, which themselves originate from the parton showering and hadronisation processes of jets;
- **Light flavour e/γ :** Real electrons coming from the decay of a light quark hadron. This includes electrons from converted photons from such decays, such as $\pi^0 \rightarrow \gamma\gamma$, which are copious in the LHC data.
- **Light flavour hadrons:** Charged hadrons (such as π^\pm) that fake the electron signature. This is the dominant background source after electron reconstruction and before electron identification are applied.

The performance of the CNN will be compared to the LLH that has a binary output, for which signal electrons include both prompt and charge-flip electrons and background electrons include all other classes.

3.2 Simulated samples used for CNN training and performance evaluation

Deep learning algorithms such as CNNs have a high learning capacity, which implies that their performance is largely dependent on the available training statistics. For this reason, several samples enriched in different types of electron signals and backgrounds are combined to maximize the electron statistics. These samples, which are listed in Table 1, form a large inclusive sample consisting of approximately 183 million reconstructed candidate electrons that are used to train the CNN and test its performance. Even though some samples are particularly enriched in a specific type of electrons, such as $Z \rightarrow e^+e^-$ for signal electrons, each sample also contains other types of electrons, such as charged hadrons faking the electron signatures for example. Therefore, each of the samples described below are allowed to populate any of the electron classes defined in Section 3.1.

The simulation [12] of the production and decay of $Z \rightarrow e^+e^-$, $Z \rightarrow \tau^+\tau^-$, Drell-Yan $\gamma^*/Z^* \rightarrow e^+e^-$, $W^+ \rightarrow e^+\nu_e$, $W^- \rightarrow e^-\bar{\nu}_e$, $W^+ \rightarrow \tau^+\nu_\tau$ and $W^- \rightarrow \tau^-\bar{\nu}_\tau$ were all performed with the POWHEG BOX v1 MC generator [13–16] at NLO accuracy. The Drell-Yan processes were generated in ten subsamples of various m_{ee} slices covering 10 – 60 GeV, 120 – 180 GeV, and 180 GeV all the way up to m_{ee} values of 5 TeV. For all of these samples, POWHEG was interfaced to PYTHIA 8.186 [17] for the modelling of the parton shower, hadronisation, and underlying event, with parameters set according to the AZNLO tune [18]. The CT10NLO parton distribution functions (PDF) set [19] was used for the hard-scattering processes, whereas the CTEQ6L1 PDF set [20] was used for the parton shower. The effect of QED final state radiation was simulated with PHOTOS 3.52 [21, 22]. The EVTGEN 1.2.0 program [23] was used to decay bottom and charm hadrons.

The production of $t\bar{t}$ events was modelled [24] using the POWHEG BOX v2 [15] generator at NLO with the NNPDF3.0NLO [25] PDF set and the hdamp parameter² set to $1.5 \times m_{\text{top}}$ [26]. The events were interfaced to PYTHIA 8.230 [27] to model the parton shower, hadronisation, and underlying event, with parameters set according to the A14 tune [28] and using the NNPDF2.3LO [29] set of PDFs. The decays of bottom and charm hadrons were performed by EVTGEN 1.6.0. During the simulation, at least one of the W bosons originating from a top quark is required to decay leptonically.

To maximize the available statistics of electron backgrounds, several high cross section two-to-two processes were simulated in individual datasets using PYTHIA 8.186 with the A14 set of tuned parameters and the NNPDF2.3LO PDF set. The simulated processes include multijet production, $qg \rightarrow q\gamma$, $q\bar{q} \rightarrow g\gamma$, W and Z boson production, as well as top quark production. No restrictions are placed on the decays of the bosons. To enrich the sample in electron backgrounds, only events where the sum of transverse energy of particles of the hard scatter (excluding muons and neutrinos) in a $\Delta\eta \times \Delta\phi$ window of 0.1×0.1 exceeds a certain threshold are selected. Three samples have been generated with lower energy thresholds of 17, 35 and 50 GeV, guaranteeing good background statistics over a wide range of p_T values. For the remainder of this note, these samples will be referred to as JF17, JF35 and JF50, respectively.

Large statistics in background electrons originating from prompt photon conversion is obtained from simulated photon+jets sample which are sliced in photon p_T . These events are generated with PYTHIA8 with the NNPDF2.3LO set of PDFs and the A14 tune.

In all of these samples, the effect of multiple interactions in the same and neighbouring bunch crossings (pileup) was modelled by overlaying the simulated hard-scattering event with inelastic proton–proton

² The hdamp parameter is a resummation damping factor and one of the parameters that controls the matching of POWHEG matrix elements to the parton shower and thus effectively regulates the high- p_T radiation against which the $t\bar{t}$ system recoils.

(pp) events generated with PYTHIA 8.186 [17] using the NNPDF2.3LO set of parton distribution functions (PDF) [29] and the A3 set of tuned parameters [30].

The number of electron candidates with $p_T > 4.5$ GeV and $|\eta| < 2.5$ for each of the process categories described above is shown in Table 1. The total number of electron candidates available is roughly 183 million in the inclusive sample, of which approximately 90% were used for the training of the CNN, and 10% for evaluating its performance. Table 2 shows the class proportions in that sample, as well as the process they come from. The majority of candidate electrons belong to the light-flavour hadron category. Signal prompt electrons are also well represented in the sample, which is important for the CNN to perform well in signal electron recognition. The sample contains at least 1 million electron candidates of each electron class, which guarantees a good identification performance across all classes.

Table 1: Number of electron candidates in the inclusive simulated sample used for training the CNN and evaluating its performance (see text for details).

Processes	Number of electron candidates
JF17, JF35, JF50	71,839,462
$t\bar{t}$ (at least one W boson $\rightarrow \ell\nu$)	63,891,541
W, Z	28,287,996
Drell-Yan $Z^*/\gamma^* \rightarrow e^+e^-$	11,823,539
γ + jet	7,406,539
Total	183,249,077

Table 2: Class composition (in %) of the inclusive sample used to train and evaluate the CNN, as well as the process which they come from.

Process \ Class	Prompt e^\pm	Charge Flip	γ Conversion	HF Decay	LF e/γ	LF Hadron
JF17, JF35, JF50	2.5×10^{-3}	9.5×10^{-5}	0.05	0.20	7.67	31.28
$t\bar{t}$ ($W \rightarrow \ell\nu$)	2.19	0.07	0.05	1.22	5.44	25.89
W, Z	10.45	0.36	0.04	0.03	1.48	3.08
$Z^*/\gamma^* \rightarrow e^+e^-$	3.93	0.21	8.6×10^{-3}	0.01	0.54	1.76
γ + jet	3.9×10^{-5}	3.8×10^{-6}	0.73	0.02	0.69	2.59
All Processes	16.57	0.64	0.88	1.49	15.82	64.60

4 Input variables to the CNN

Three types of features are used as input to the CNN: high-level variables (which are used also in the LLH and DNN), tracking variables and the calorimeter images. Each type of features is described in Sections 4.2-4.4. These variables are computed once the electron candidates are reconstructed, which procedure is first briefly described in Section 4.1.

4.1 Electron reconstruction

Electron candidates are reconstructed by matching a cluster of energy deposits in the calorimeter with at least one inner detector track. The clusters are built using the topo-cluster reconstruction algorithm [31, 32]. To reduce the amount of topo-clusters reconstructed from pile-up, only clusters with at least half of their energy deposited in the electromagnetic calorimeter are considered for the electron reconstruction. These topo-clusters are matched to tracks that are re-fitted to account for bremsstrahlung [33]. To also capture possible energy deposits from bremsstrahlung, super-clusters are formed which can contain multiple topo-clusters. Electron candidates with $p_T > 4.5$ GeV and $|\eta| < 2.5$ are retained for further analysis. More details on electron reconstruction can be found in Refs. [1, 2].

4.2 High-level variables

The CNN uses several high-level variables (HLV) which are listed in Table 3 for the variables relying solely on calorimetry information, and Table 4 for all other variables. We note that, even though the HLV could in principle be reconstructed by the neural network using the information from the additional tracks and the calorimeter images, we observe that the HLV are crucial to help the CNN converge and find a good global minimum during its training, and generally obtain good performance. The HLV list is the same as for the LLH and DNN with a couple additions. The charge of the best matched track times the absolute value of its transverse impact parameter ($q \times d_0 / \sigma(d_0)$), as well as the average charge of all tracks matched to the electron weighted by their number of hits in the SCT detector ($\langle q \rangle_{\text{SCTw}}$), are both used by the CNN while not being included for the DNN and LLH algorithms. These variables have been developed for the identification of charge-flip electrons [1]. Similarly to the DNN, the E_T and η of the electron candidates are used as input to parametrise the CNN as a function of these variables, while they are not used by the LLH.

It is to be noted that the tracking variables in Table 4 are those corresponding to the main electron track identified by the electron reconstruction algorithm, while the features of additional tracks associated to an electron candidate are also considered separately by the CNN as discussed in the following section.

4.3 Additional track variables

The ATLAS electron reconstruction algorithm can associate more than one track to an electron candidate. These additional tracks contain useful information since, for example, the average number of tracks for signal electrons in the inclusive sample described in Section 3 is 1.3, while it is 2.8 for background classes. The 13 features listed in Table 5 are considered by the CNN for up to five matched tracks, which are ordered in increasing value of their distance to the reconstructed electron candidate position in $\Delta\eta \times \Delta\phi$.

4.4 Calorimeter images

More optimal electron identification performance can be achieved by using the two-dimensional images of the energy deposited in each of the calorimeter layers around an electron candidate. There is the technical complication in creating such images that the number of calorimeter layers that an electron candidate will cross varies as a function of $|\eta|$ (e.g. because the barrel and endcap calorimeter overlap in the transition region $1.37 < |\eta| < 1.52$), and that the calorimeter granularity is not the same for every layer

Table 3: Calorimeter input high-level variables to the CNN.

Type	Description	Symbol
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster	R_{had1}
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster	R_{had}
Third layer of EM calorimeter	Ratio of the energy in the third layer to the total energy in the EM calorimeter. Due to known mismodelling at high $ \eta $, this variable is set to a default value for candidates with $ \eta > 2.01$	f_3
Second layer of EM calorimeter	Lateral shower width, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	$w_{\eta 2}$
	Ratio of the energy in 3×3 cells over the energy in 3×7 cells centred at the electron cluster position	R_ϕ
	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centred at the electron cluster position	R_η
First layer of EM calorimeter	Shower width, $\sqrt{(\sum E_i (i - i_{\text{max}})^2)/(\sum E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η , and i_{max} is the index of the highest-energy strip	w_{stot}
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies	E_{ratio}
	Ratio of the energy in the first layer to the total energy in the EM calorimeter	f_1

and η region. To tackle the first issue, images of every layer of the calorimeters are produced regardless of the pseudorapidity of the electron. So, for example, images of the endcap calorimeter are created for electrons produced in the barrel region, and are filled with zeros. All images are produced with a granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, which corresponds to the granularity of the second layer of the electromagnetic calorimeter that collects most of the electron's energy. Images for calorimeter layers with a coarser granularity are produced by distributing the deposited energy into images of granularity $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. For example, for a Tile calorimeter cell of granularity $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, its deposited energy is distributed equally between the 16 "pixels" of granularity $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ that corresponds to the area of that Tile calorimeter cell. There is the exception of the first calorimeter layer that reaches even finer granularities that depend on $|\eta|$ (see Section 2). The images for this layer are all produced with the finest available granularity for that layer: $0.025/8$. Each calorimeter image is centered on the electron candidate calorimeter cluster position. The images are 0.175 wide in η and 0.275 wide in ϕ for all calorimeter layers, which corresponds to the size of 7 cells in η and 11 cells in ϕ for the layer 2 of the EM calorimeter. A cell is included in a calorimeter image if (and only if) its center is located within these ranges, and its entire energy is used. Figure 1 shows the calorimeter images averaged over all electron candidates with $|\eta| < 1.3$ in the barrel section. The corresponding images in the endcap region $1.85 < |\eta| < 2.5$ are shown in Figure 2. The color of these images represents the mean energy deposited divided by the electron energy (in %). Only calorimeter layers with non-zero deposited energies in each η

Table 4: Track, track-cluster matching and kinematic input high-level variables to the CNN.

Type	Description	Symbol
Track quality	Number of hits in the innermost pixel layer	n_{Blayer}
	Number of hits in the pixel detector	n_{Pixel}
	Number of hits in the SCT detector	n_{SCT}
	Number of degrees of freedom in the track fit	n_{dof}
	Transverse impact parameter relative to the beamline	d_0
	Significance of transverse impact parameter defined as the ratio of d_0 to its uncertainty	$d_0/\sigma(d_0)$
	Momentum lost by the track between the perigee and the last measurement point divided by the momentum at perigee	$\Delta p/p$
	Charge of the track times the significance of the transverse impact parameter	$q \times d_0/\sigma(d_0)$
	Average charge of all tracks matched to the electron weighted by their number of hits in the SCT detector	$\langle q \rangle_{\text{SCTw}}$
TRT	Likelihood probability based on transition radiation in the TRT. This variable is set to a default value for candidates with $\eta > 2.01$ due to the limited coverage of the TRT.	$e\text{ProbabilityHT}$
Track-cluster matching	$\Delta\eta$ between the cluster position in the first layer and the extrapolated track	$\Delta\eta_1$
	$\Delta\phi$ between the cluster position in the second layer of the EM calorimeter and the momentum-rescaled track, extrapolated from the perigee, times the charge q	$\Delta\phi_{\text{res}}$
	Ratio of the cluster energy to the track momentum	E/p
	Ratio of the cluster transverse energy to the track p_T	E_T/p_T
	Number of tracks matching the electron candidate	n_{tracks}
Kinematics	Transverse energy of the electron measured by the calorimeter system. This variable is not used for discrimination purposes but for the CNN to infer the E_T dependence of other variables.	E_T
	Pseudorapidity of the electron as measured by the calorimeter system. This variable is not used for discrimination purposes but for the CNN to infer the η dependence of other variables.	η

region are shown. As expected, significant differences in the transverse and longitudinal developments of the showers, as well as the activity neighboring the electron candidate, can be observed between the different signal and background classes. It is to be noted that every calorimeter layer has been found to contribute non-negligibly to the CNN performance, and thus all of them are considered as inputs to the CNN.

Table 5: CNN input variables for each track j which is matched to an electron candidate, up-to a maximum of five tracks. For the computation of $\Delta\eta$ and $\Delta\phi$, the electron candidate position is determined to be the one of the best matched track to the electron candidate.

Type	Description	Symbol
Matching variables	Ratio of the momentum of track j to the energy of the electron candidate	p_j/E
	$\Delta\eta$ between the track j and the electron candidate position	$\Delta\eta_j$
	$\Delta\phi$ between the track j and the electron candidate position	$\Delta\phi_j$
Number of hits	Number of hits in the pixel detector	n_{Pixel}^j
	Number of hits in the SCT detector	n_{SCT}^j
	Number of hits in the TRT detector	n_{TRT}^j
Track parameters and fit quality	Transverse impact parameter relative to the beamline	d_0^j
	Uncertainty on d_0	$\sigma(d_0)_j$
	Longitudinal impact parameter relative to the beamline	z_0^j
	Charge of the track	q_j
	χ^2 of the track fit	χ_j^2
	Number of degrees of freedom of the track fit	ndof_j
	Matched vertex index number	vtx_j

5 Description of the CNN

5.1 Preparation of the Training Sample

The distributions of E_T and η for a subsample of 18.8 million electron candidates of the samples described in Section 3 are shown in Figure 3. The structures in the E_T distribution is due to the fact that some samples are sliced kinematically, such as the Drell-Yan samples in m_{ee} and the γ +jets samples in photon p_T . The η distributions of most electron classes is maximal around $\eta \approx 0$ and drops slowly with increasing $|\eta|$, except for the charge-flip, light flavor e/γ and photon conversion classes that involve material effects, and thus are either flat or increasing as a function of $|\eta|$. In order for the CNN to be as physics-independent as possible, the normalisation and shapes of the E_T and η distributions of each background class are reweighed in two dimensions to match the ones of signal electrons. In this manner, the CNN will not attempt to distinguish signal versus background based on the values of E_T and η of the electron candidate. In addition, reweighing the normalisation of each class to the same value allows to mitigate significant differences in class ratios, such as those observed in Figure 3, which have proven to be detrimental for the discrimination performance of less prevalent classes such as charge-flip and photon-converted electrons. In the same procedure, the background to signal class ratio can be adjusted to optimize the performance. This ratio was treated as an hyper-parameter and found to be optimal around a value of 4 for the inclusive sample.

5.2 Preprocessing

Before being fed to the neural network, high-level variables are transformed via the QuantileTransformer class available in the scikit-learn Python library [34], which allows each variable to follow a standard

ATLAS Simulation Preliminary ; $\sqrt{s} = 13 \text{ TeV}$; $|\eta| < 1.3$

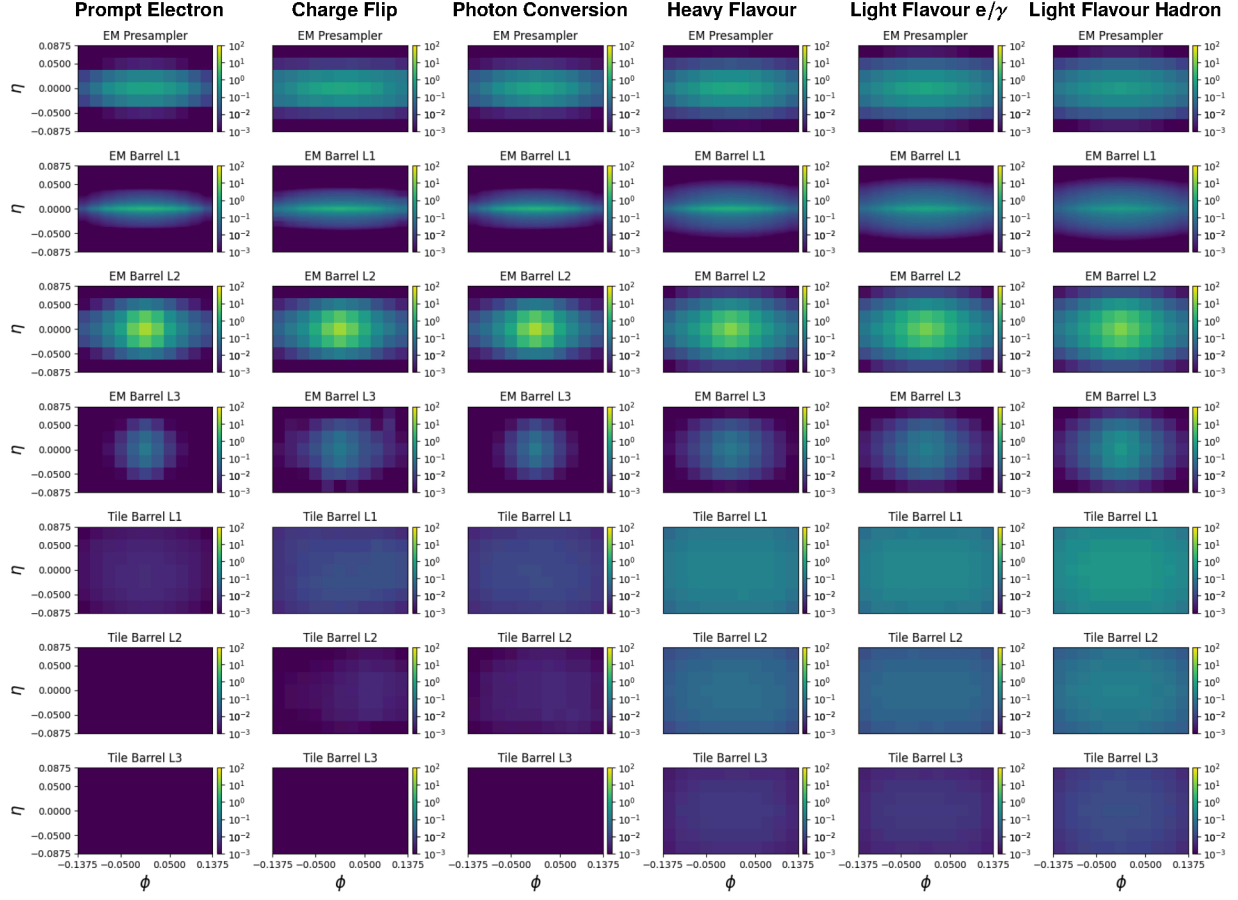


Figure 1: Mean energy deposits images for electron candidates with $|\eta| < 1.3$ in a given cell divided by the total energy of the electron candidate (in %). Each column corresponds to the different electron classes. The top row corresponds to the presampler layer which is followed by the three layers of the barrel electromagnetic calorimeter. Then the last three rows show the hadronic tile calorimeter layers. The images of each layer are treated with a granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, except the first layer of the EM calorimeter for which it is $\Delta\eta \times \Delta\phi = 0.025/8 \times 0.025$.

normal distribution. This non-linear transform is applied independently to each feature and allows the learning algorithm to perform slightly better. It also tends to spread out most frequent values and reduce the impact of outliers. No transformation is applied for the calorimeter images and additional tracks features.

5.3 Network Architecture

Figure 4 shows the network architecture of the CNN model trained for electron identification. Three different sub-CNNs are deployed. The first one processes the image extracted from the first electromagnetic calorimeter layer, which has a finer granularity of 56×11 in $\eta \times \phi$. The second sub-CNN processes all of the calorimeter layers each using an $\eta \times \phi = 7 \times 11$ granularity, including the first EM calorimeter layer down-scaled in resolution. The reason to also include the layer 1 of the EM calorimeter layer in the second sub-CNN is to capture the longitudinal development of the shower of the electron candidate.

ATLAS Simulation Preliminary ; $\sqrt{s} = 13$ TeV ; $1.85 < |\eta| < 2.5$

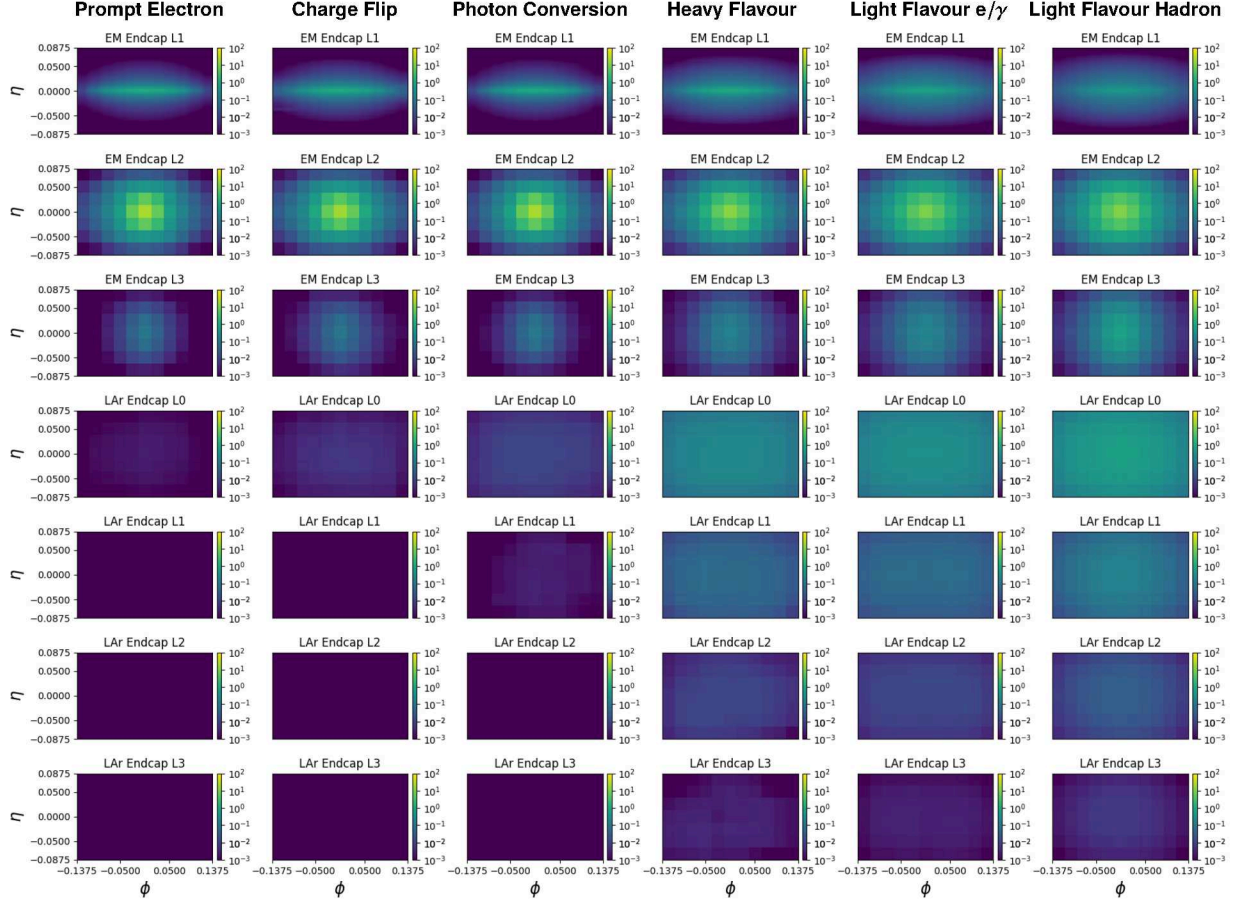


Figure 2: Mean energy deposits images for electrons candidates with $1.85 < |\eta| < 2.5$ in a given cell divided by the total energy of the electron candidate (in %). Each column corresponds to the different electron classes. The top rows correspond to the three layers of the endcap electromagnetic calorimeter, followed by the four rows of the hadronic endcap calorimeter layers. The images of each layer are treated with a granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, except the first layer of the EM calorimeter for which it is $\Delta\eta \times \Delta\phi = 0.025/8 \times 0.025$.

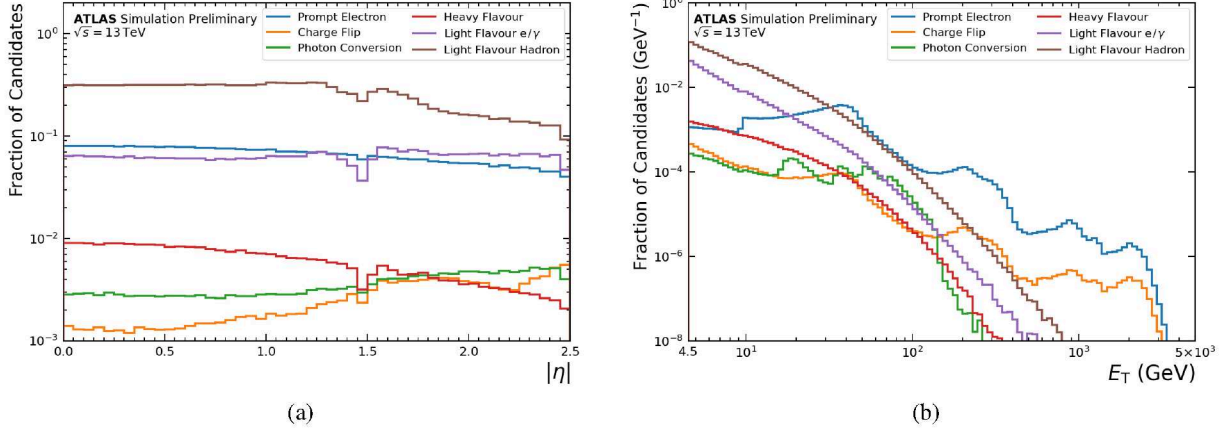


Figure 3: Pseudorapidity and transverse energy of the approximately 18.8 million electron candidates used to validate the CNN performance.

The calorimeter images from the different layers are treated as different channels in the second sub-CNN (like red-blue-green in a standard colored image-processing CNN). The last sub-CNN is used to process the information of the tracks associated with the electron object using a 1×1 kernel. The outputs are flattened, concatenated and fed to fully-connected neural network layers, together with the HLV input features. This fully connected network outputs the probabilities corresponding to the signal or background classes introduced in Section 3.

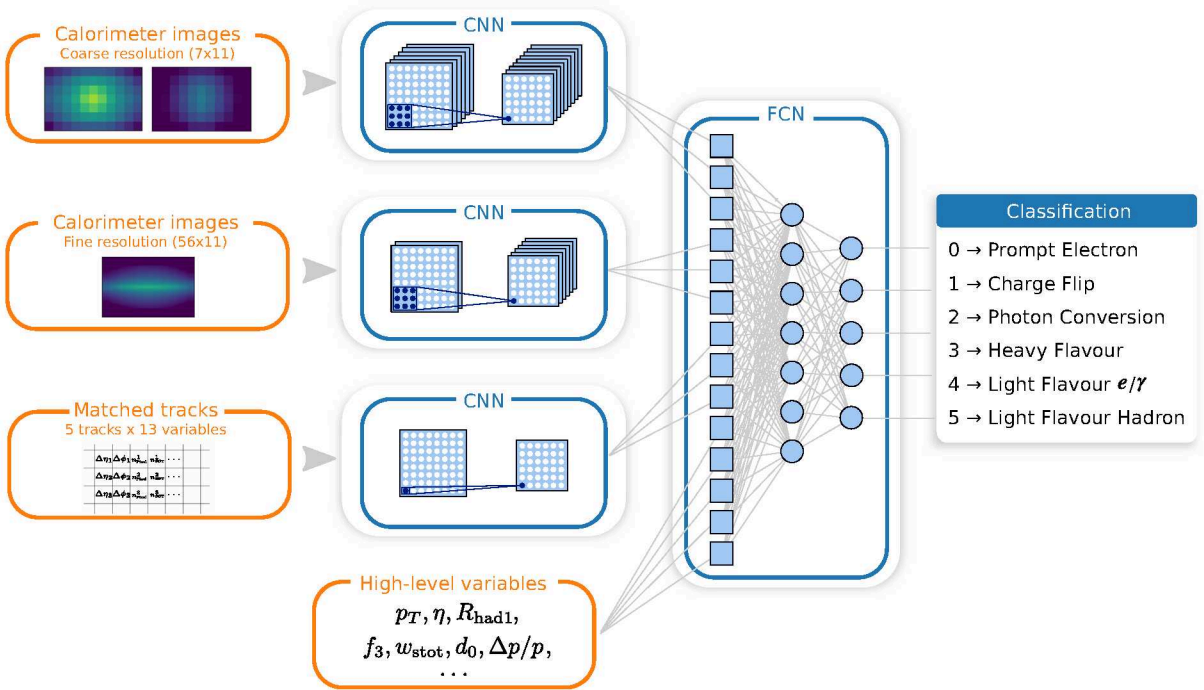


Figure 4: Global neural network architecture.

Table 6 describes the hyper-parameters that have been optimized for each of the four network components shown in Figure 4. Each of the three sub-CNNs has two consecutive convolution layers for which the

number of neurons are constrained by specific kernel and pooling sizes (respectively used to perform the convolutions and subsequently reduce their resulting dimensionality). Each of these two layers has a certain number of feature maps chosen to extract salient properties of the images. The outputs of all sub-CNNs are flattened and concatenated along the values of high-level variables and then connected to a fully-connected network, which in turn consists of two hidden layers of 100 neurons each, followed by an output layer of six neurons (i.e. the number of classes).

Table 6: Hyper-parameters for each of the components of the neural network.

Network Component	Kernel Sizes	Pooling Sizes	Feature Maps	Hidden Neurons
Coarse images (7×11)	(3,5), (3,5)	(1,1), (1,1)	100, 100	-
Fine images (56×11)	(3,5), (3,5)	(4,1), (2,1)	100, 100	-
Matched tracks (5×13)	(1,1), (1,1)	(1,1), (1,1)	200, 200	-
High-level variables (26)	-	-	-	100, 100

The network is trained using `TENSORFLOW` [35]. Each hidden layer is processed through the ReLU activation function whereas the Softmax activation function is used to normalize the output layer to a probability distribution over predicted output classes. The gradient descent is performed with the `ADAM` optimiser [36] and uses a variable adaptive learning rate that progressively gets smaller as the minimum of the loss function gets closer. The categorical cross-entropy is employed as the loss function and the accuracy is used as the metric to assess training convergence. The algorithm regularly writes checkpoints to save the weights of the best epoch. The training is terminated after five consecutive epochs for which the metric did not improve from the previous epoch, thus signalling the convergence has reached a plateau and making the algorithm retrieve the weights of the best epoch. Throughout the training, L2 [37] and Dropout [38] regularisation are used in order to minimize the risk of overfitting.

5.4 Definition of the Final Discriminant

The CNN outputs for each electron candidate a vector normalised to unity with six entries, each representing the estimated probability that the candidate belongs to one of the six classes defined in Section 3. Although in principle the probability for an electron candidate to be signal p_{sig} could be used as the final discriminant, in practice a more powerful discriminant [39] can be formed by taking the ratio of the probabilities to be signal or background:

$$\mathcal{D} = \frac{\sum w_{\text{sig}} p_{\text{sig}}}{\sum w_{\text{bkg}} p_{\text{bkg}}}, \quad (1)$$

where the sums on the numerator and denominator are on the signal and background classes, respectively. The parameters w_{sig} and w_{bkg} are adjustable weights, of which the sums are constrained to unity. For most use cases, charge-flip electrons are included as signal (in addition to prompt electrons) because many physics analyses do not have requirements on the electron charge. For the CNN, the discriminant \mathcal{D} is transformed as:

$$\mathcal{D} \rightarrow \mathcal{D}/(1 + \mathcal{D}), \quad (2)$$

such that the new discriminant is bounded between 0 and 1, which does not alter the performance. As demonstrated in the DNN note [3], the chosen values for the weights w_{sig} and w_{bkg} affect significantly the electron identification performance. For instance, the background rejection can be increased against a

specific background class by increasing the value of w_{bkg} for that class. The CNN is found to perform approximately optimally against the combined background of a given sample when w_{sig} and w_{bkg} are set to the actual fractions in which each class occurs in that sample. It will be shown in Section 6.2 that, other than performing electron identification, the CNN output can be used to estimate the true class fractions of a given sample.

6 Performance

6.1 Analysis of the performance and comparison with the Likelihood algorithm

The performance of the CNN is estimated in a sample that is statistically independent to the one used for training the neural network. It consists of 10% of the inclusive sample described in Section 3, with the same electron class proportions, representing approximately 18.8 million electron candidates. Given that the CNN achieves high background rejection, especially for the light-flavour classes, it is important to have a high-statistics evaluation sample to be able to properly characterize its performance, in particular differentially as a function of E_T , η and pile-up activity, as presented below. There is no kinematic reweighting applied to the validation sample, such that Figure 3 reflects the actual E_T and η distributions of the candidate electrons in that sample. Standard electron quality requirements are applied to that sample. The number of pixel and silicon (pixel+SCT) hits on the main electron track are required to be at least two and seven, respectively. In addition, the same ambiguity resolution cut than the LLH Loose and Medium working point is applied to distinguish objects that are reconstructed as both electrons and photons. These quality requirements result in an evaluation sample of 17.5 million electron candidates.

The shape of the CNN discriminant of Equation 2 is shown in Figure 5, for which the parameters w_{sig} and w_{bkg} of Equation 1 were set to the true class proportions of the combined sample (see Table 2). Excellent separation is observed, in particular for the light-flavour hadron class, which is the only background class that is not prominently composed of real (non-prompt) electrons. Still, it is important to achieve good rejection for that background class since it is by far the most prominently produced in the LHC data. The other background classes tend to be more difficult to reject than light-flavour hadrons, but good discrimination can still be achieved.

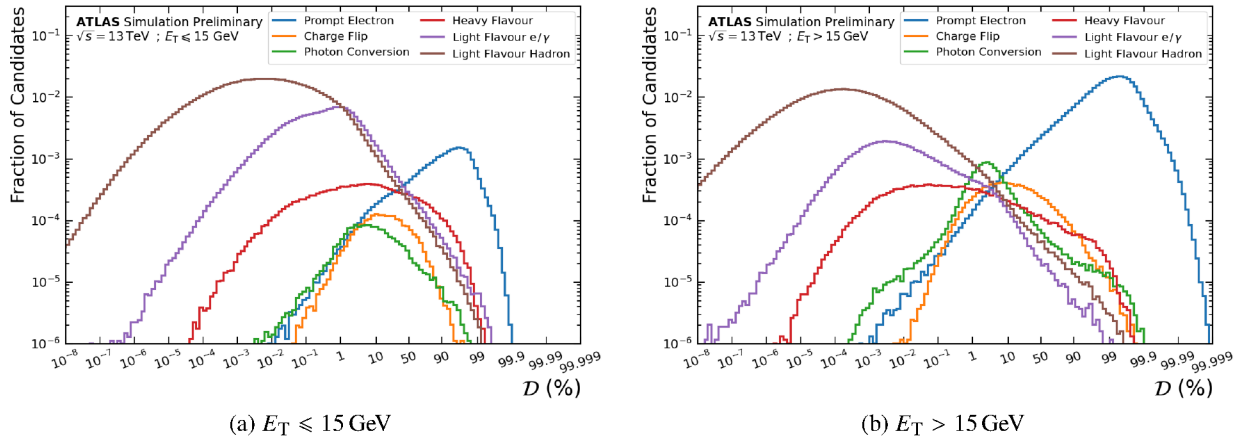
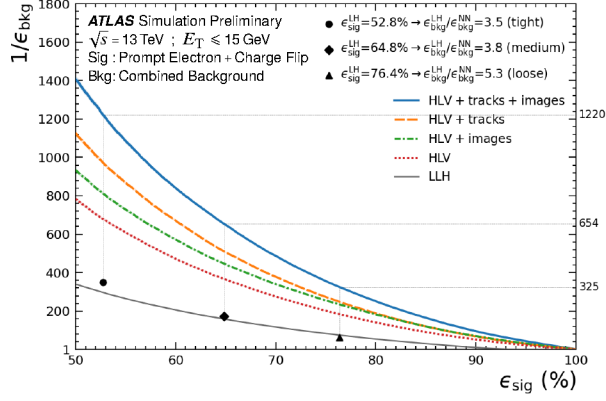


Figure 5: The CNN discriminant for the various electron classes for (a) $4.5 \text{ GeV} < E_T \leq 15 \text{ GeV}$ and (b) $E_T > 15 \text{ GeV}$

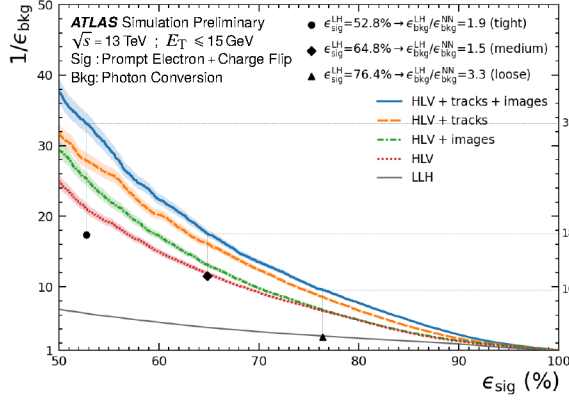
The performance of the CNN can be further characterized from the Receiver Operating Characteristics (ROC) curves that show the signal efficiency versus the background rejection (i.e. the inverse of the background efficiency $1/\epsilon_{\text{bkg}}$). The ROC curves are obtained by varying the cut value on the discriminant and computing at each value the signal and background efficiencies. Figures 6 and 7 shows the ROC curves for individual classes for the case where charge-flip electrons are included in the signal (the most popular use-case in ATLAS) for electron candidates with $4.5 \text{ GeV} < E_T \leq 15 \text{ GeV}$ and $E_T > 15 \text{ GeV}$, respectively (inclusively in η). In these figures, the performance of the CNN using all types of inputs, i.e. HLV, additional tracks and the calorimeter images, are shown with the solid blue lines. The dotted red, dash-dotted green and dashed orange lines also show the performance of the algorithms when only the HLV, HLV+images and HLV+track inputs are used, respectively, to illustrate which component of the CNN is most powerful against which type of background. The corresponding ROC curves for the LLH are shown as solid black lines, as well as the Loose, Medium and Tight working points that are shown as black markers on these figures. The working point markers do not necessarily overlap with the LLH ROC curves because the cut on the LLH discriminant varies as a function of p_T and η in the definition of the working points. Furthermore, additional requirements are applied for the LLH working points, such as requiring a hit on the innermost pixel layer for the Medium and Tight working points, which is particularly useful to reject the photon conversion background.

Large improvements with respect to the LLH can be obtained for all electron classes. For example, improvement factors in background rejection of 4.3 and 6.9 are observed against the light flavour e/γ and hadrons, respectively, for the medium working points for $E_T > 15 \text{ GeV}$. These are the easiest classes to classify, and also the ones for which the calorimeter images are the most distinctive from signal electrons. Significant improvements over the LLH are also observed for photon conversion and heavy-flavour decays, as well as against the combined background of the evaluation sample as shown in Figures 6(a) and 7(a). One exception is the performance against photon conversion at low p_T where the performance is comparable to the LLH, which is the class with less statistics in the training sample in that p_T range. Another conclusion that can be reached from these figures is that, in addition to the HLV that are common to the LLH, DNN and CNN, the calorimeter images and the additional tracks have different importance against different types of electron backgrounds. The calorimeter images play a more prominent role in rejecting light flavour e/γ and hadrons (in particular at low p_T), presumably because of their more prominent surrounding hadronic activity, in particular for light flavour hadrons that feature hadronic showers instead of electromagnetic showers in the calorimeter. The calorimeter images are not as powerful at high p_T , which is likely due to the fact that they are the component of the CNN that requires the most statistics to perform adequately, and the two light flavour categories are much more prominent at low than at high p_T in the training sample, as illustrated in Figure 3. For prompt photon conversion and heavy flavour electrons, the additional track variables are particularly important presumably because, like signal electrons, they feature electromagnetic showers, but they tend to have more than one tracks matching a candidate electron. In general, all features are required to obtain optimal performance.

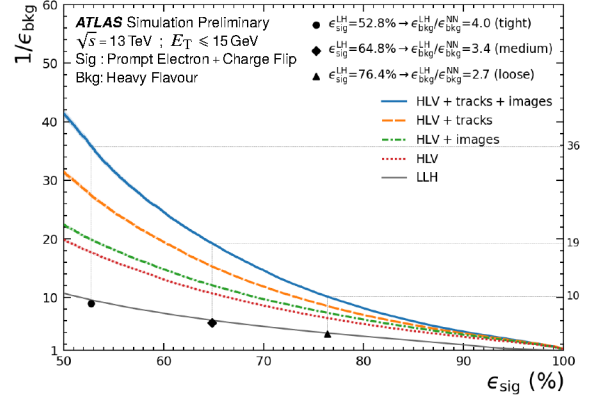
Similar ROC curves are shown in Figures 8 ($4.5 \text{ GeV} < E_T \leq 15 \text{ GeV}$) and 9 ($E_T > 15 \text{ GeV}$) for the case where charge-flip electrons are included in the background, as is the case for analyses featuring same-sign electron pairs for example. Specifically for the ROC curves of performance against charge-flip electrons (Figures 8(b) and 9(b)), the performance of the charge-flip electron identifier (ECIDS) [1], which is based on a boosted decision tree, is shown when applied on top of the LLH, as it is in most ATLAS analyses that suffer from charge-flip electron backgrounds. For $4.5 \text{ GeV} < E_T \leq 15 \text{ GeV}$, which is not a common use-case for same-sign lepton pair analyses at the LHC, the CNN is performing comparably to the LLH+ECIDS. For $E_T > 15 \text{ GeV}$, large improvements in background rejection are obtained over the LLH+ECIDS combination that range from factors of 3.8 to 11.1 depending on the signal efficiency working



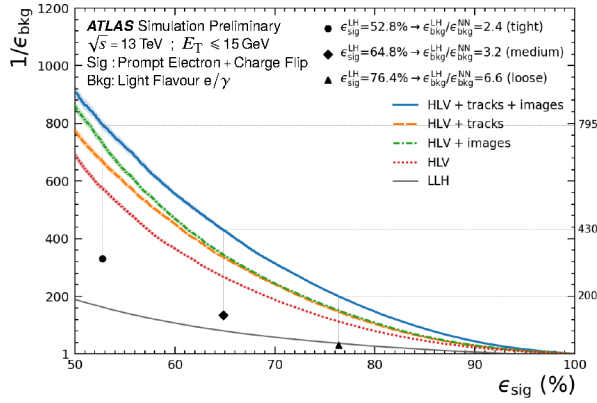
(a) Combined Background



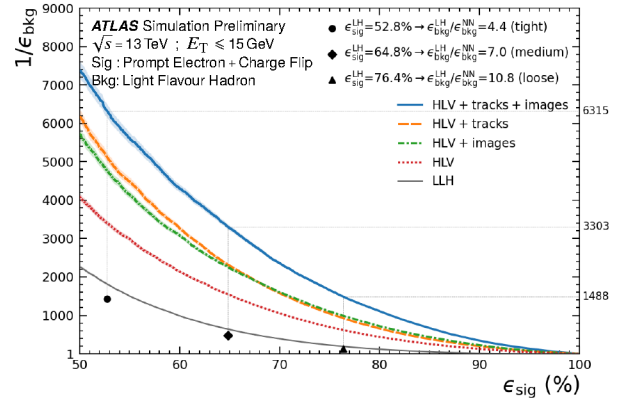
(b) Photon Conversion



(c) Heavy Flavour

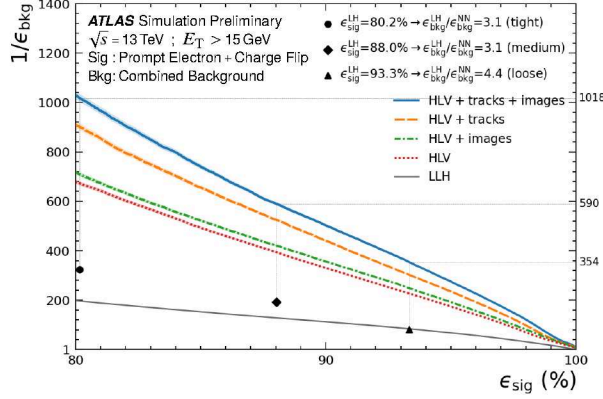


(d) Light Flavour e/γ

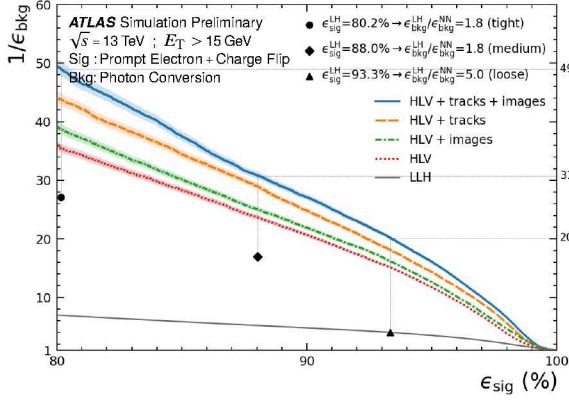


(e) Light Flavour Hadron

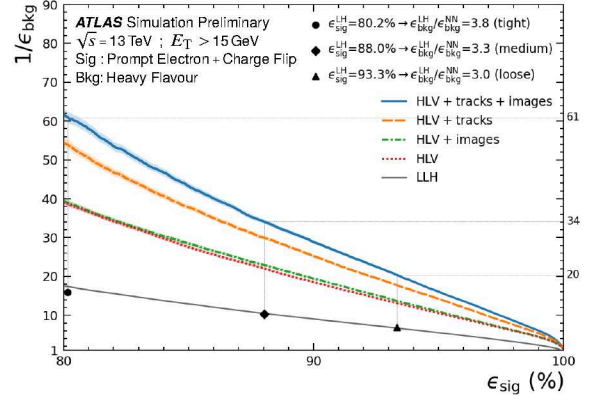
Figure 6: ROC curves for electrons with $4.5 \text{ GeV} < E_T \leq 15 \text{ GeV}$ for the case where signal electrons include both prompt and charge-flip electrons. The dotted red, dash-dotted green and dashed orange lines show the performance of the CNN when only the HLV, HLV+images and HLV+track inputs are used, respectively. The bands surrounding these lines shows the statistical uncertainty due to the limited statistics of the validation sample. The performance of the LLH for the Loose, Medium and Tight working points are shown as black markers, and the ROC curve of the LLH discriminant is shown as the solid black line.



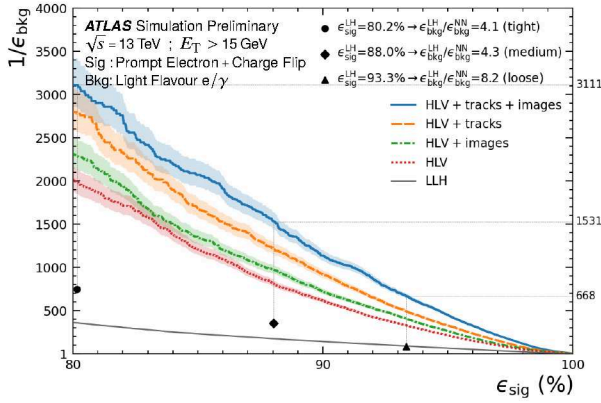
(a) Combined Background



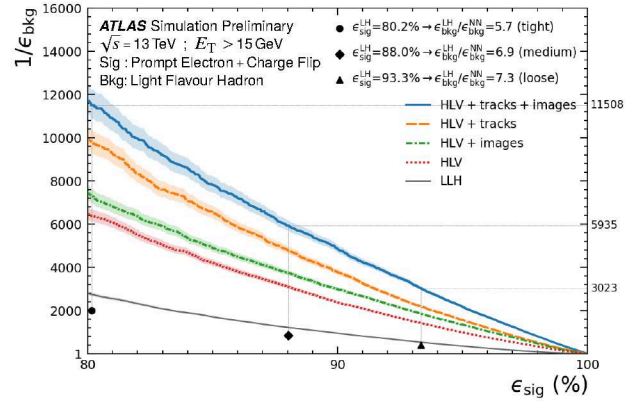
(b) Photon Conversion



(c) Heavy Flavour



(d) Light Flavour e/γ



(e) Light Flavour Hadron

Figure 7: ROC curves for electrons with $E_T > 15$ GeV for the case where signal electrons include both prompt and charge-flip electrons. The dotted red, dash-dotted green and dashed orange lines show the performance of the CNN when only the HLV, HLV+images and HLV+track inputs are used, respectively. The bands surrounding these lines shows the statistical uncertainty due to the limited statistics of the validation sample. The performance of the LLH for the Loose, Medium and Tight working points are shown as black markers, and the ROC curve of the LLH discriminant is shown as the solid black line.

point. In all of the ROC curves of Figures 6(a)- 9(a) that show the signal efficiency versus the combined background rejection, the CNN is capable of achieving a comparable or better background rejection than the LLH Tight working point for a signal efficiency that correspond to the Loose LLH working point.

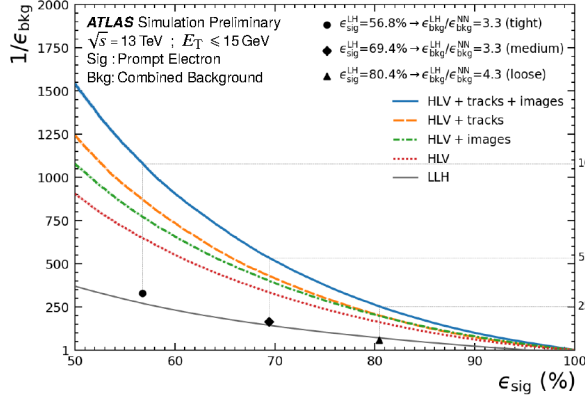
The performance of the CNN (HLV+tracks+images) is further scrutinized in Figure 10 that shows the ratios of background rejection of the CNN over the LLH as a function of E_T for signal efficiencies that match the ones of the LLH Loose, Medium and Tight working points in each bin individually. In addition, the ratios of background rejection is shown with respect to the CNN where only the HLV are used as input, which corresponds to a fully-connected neural network (FCN) that is similar to the DNN. This FCN serves to illustrate the improved performance that are due to the usage of the calorimeter images and the additional tracks. In this figure, charge-flip electrons are considered background to be able to visualize the identification performance for all types of electrons. The CNN performs better than both the LLH and FCN in all E_T regions and for all signal efficiency working points. The relative improvements as a function of both E_T and η for a signal efficiency corresponding to the LLH Loose working point (applied individually in each bin) are shown in Figure 11 for the combined background of the validation sample. Significant improvements with respect to the LLH and the FCN can be found for all values of E_T and η . Particularly large improvements are observed in the crack region ($1.37 < |\eta| < 1.52$).

Since the calorimeter image window is relatively wide around the electron candidate, it is a fair concern whether the observed improvements would lessen significantly if isolation criteria are also applied, as is usually the case in physics analyses. This is tested by applying isolation criteria to all electron candidates where the amount of transverse energy inside a cone of radius $\Delta R = 0.2$ in $\eta \times \phi$ around the electron is required to be less than 20% of of the electron E_T . Similarly, the scalar sum of the p_T of tracks located inside a cone of $\Delta R = 0.3$ is required to be less than 15% of the electron E_T . Figure 12 shows the same results as Figure 11 with in addition these isolation criteria applied around the electron candidates that are typical of searches for supersymmetry in ATLAS. Very similar improvements of the CNN with respect to the LLH or FCN are observed, whether isolation is applied or not.

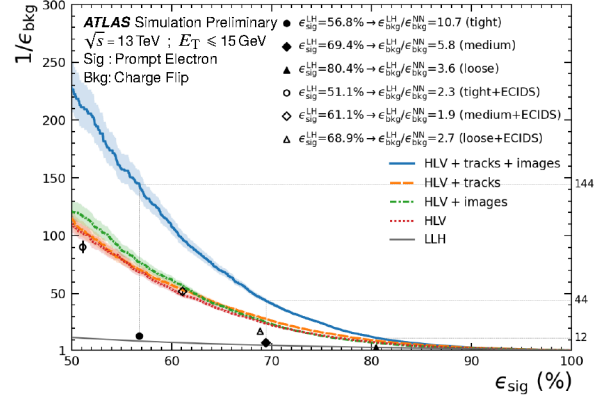
The background rejection and signal efficiency of the CNN and LLH are studied as a function of p_T and η in Figures 13 and 14, respectively. In these figures, the CNN signal efficiency is fixed to match the one of the LLH in each individual bin for the Loose (upper plots), Loose+isolation (middle plots, same isolation criteria as Figure 12) and Tight (bottom plots) LLH working points. The signal efficiency for the CNN thus follows exactly the trends of the LLH, but with a significantly improved background rejection. The robustness of the CNN performance as a function of pile-up activity is put to the test in Figure 15 that is analogous to Figures 13-14, but this time as a function of the average number of interactions per bunch crossing ($\langle\mu\rangle$). Again, the signal efficiency of the CNN follows closely the behavior of the LLH, but with much improved background rejection.

6.2 Determination of the population of electron classes in a sample using the CNN

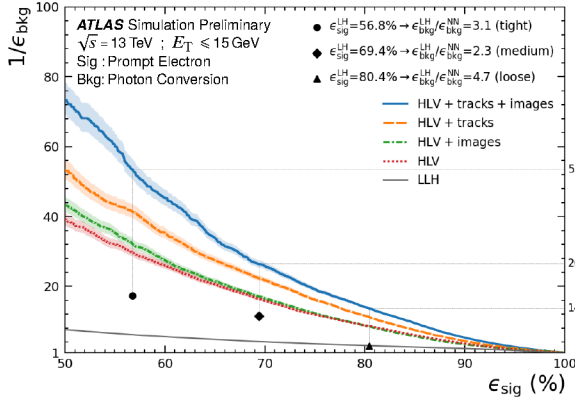
The CNN returns as output a vector of the probabilities for each electron to belong to one of the six classes described in Section 3. This opens up the possibility for the CNN to be used to determine the electron class fractions in a sample. This is achieved by assuming that the true class of each electron candidate corresponds to the one with the highest value in the predicted probability vector. When applied to an entire sample of electrons, this method provides the predicted amount of electron in each class in that sample. Figure 16(a) shows the true class ratios for a few representative simulated samples and Figure 16(b) shows the difference between the predicted and the true class ratios as a function of the true electron class on the



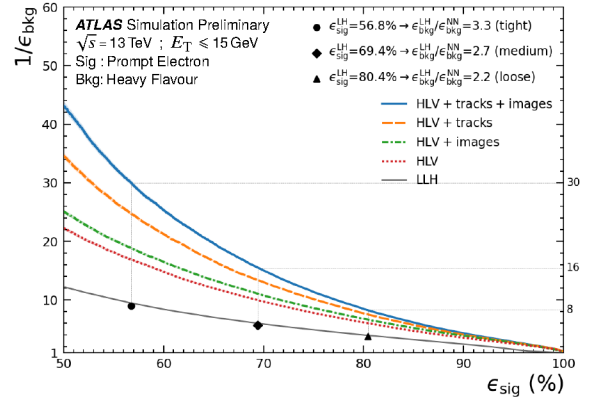
(a) Combined Background



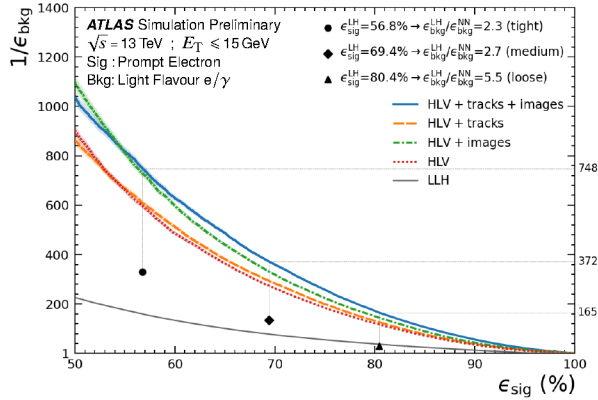
(b) Charge Flip



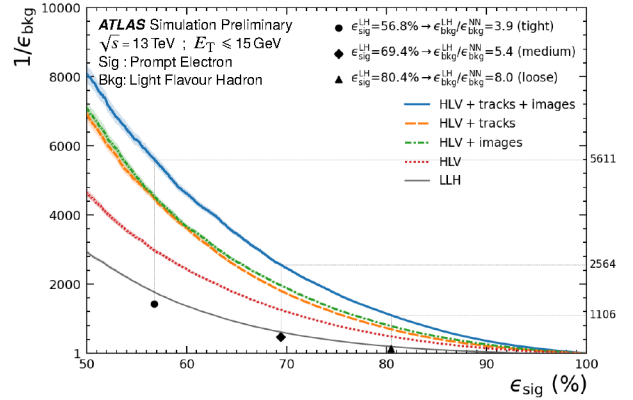
(c) Photon Conversion



(d) Heavy Flavour

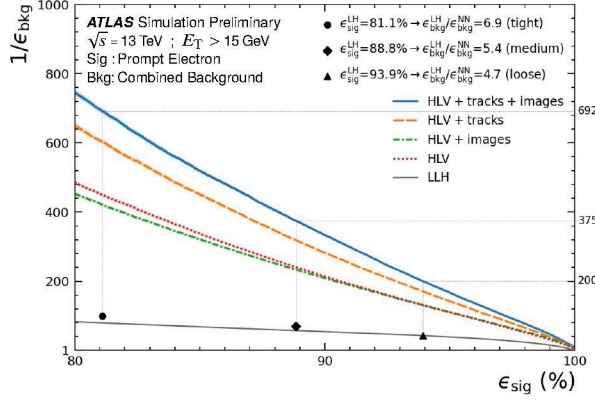


(e) Light Flavour e/γ

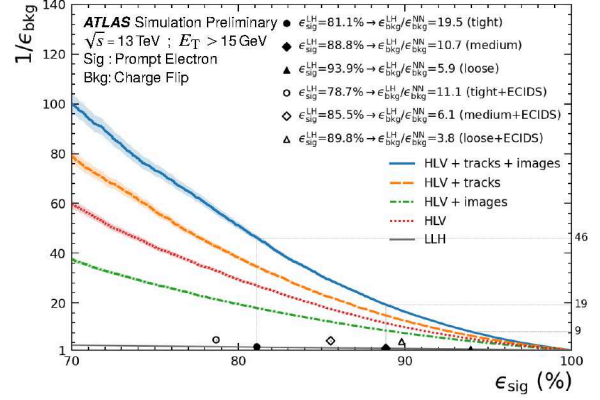


(f) Light Flavour Hadron

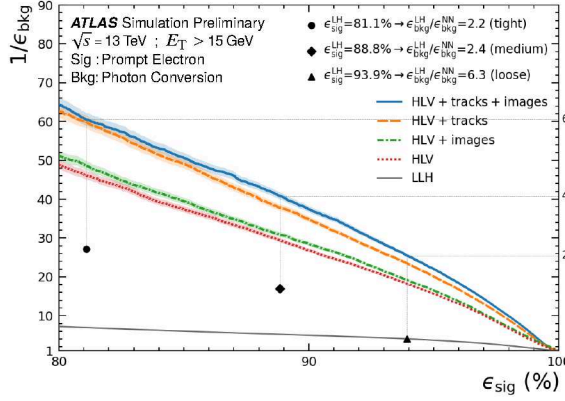
Figure 8: ROC curves for electrons with $4.5 \text{ GeV} < E_T \leq 15 \text{ GeV}$ for the case where signal electrons include only prompt electrons, and for which charge-flip electrons are classified as background. The dotted red, dash-dotted green and dashed orange lines show the performance of the CNN when only the HLV, HLV+images and HLV+track inputs are used, respectively. The bands surrounding these lines show the statistical uncertainty due to the limited statistics of the validation sample. The performance of the LLH for the Loose, Medium and Tight working points are shown as black markers, and the ROC curve of the LLH discriminant is shown as the solid black line.



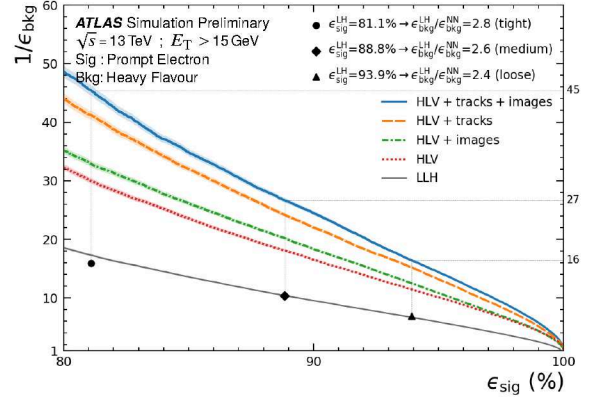
(a) Combined Background



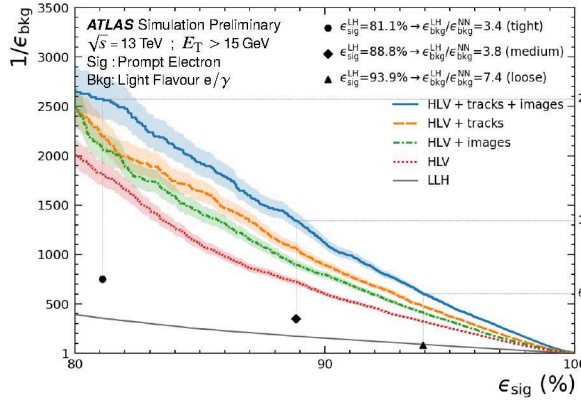
(b) Charge Flip



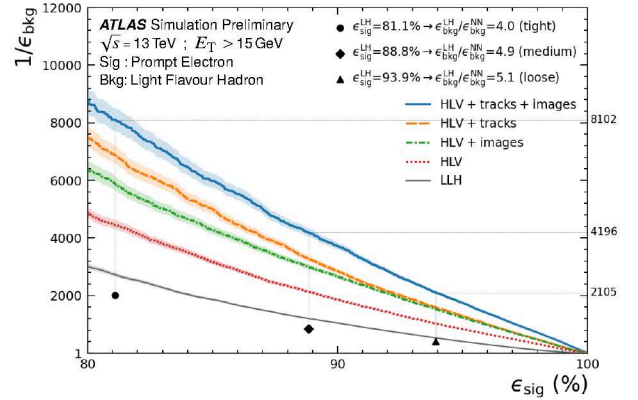
(c) Photon Conversion



(d) Heavy Flavour



(e) Light Flavour e/γ



(f) Light Flavour Hadron

Figure 9: ROC curves for electrons with $E_T > 15$ GeV for the case where signal electrons include only prompt electrons, and for which charge-flip electrons are classified as background. The dotted red, dash-dotted green and dashed orange lines show the performance of the CNN when only the HLV, HLV+images and HLV+track inputs are used, respectively. The bands surrounding these lines show the statistical uncertainty due to the limited statistics of the validation sample. The performance of the LLH for the Loose, Medium and Tight working points are shown as black markers, and the ROC curve of the LLH discriminant is shown as the solid black line.

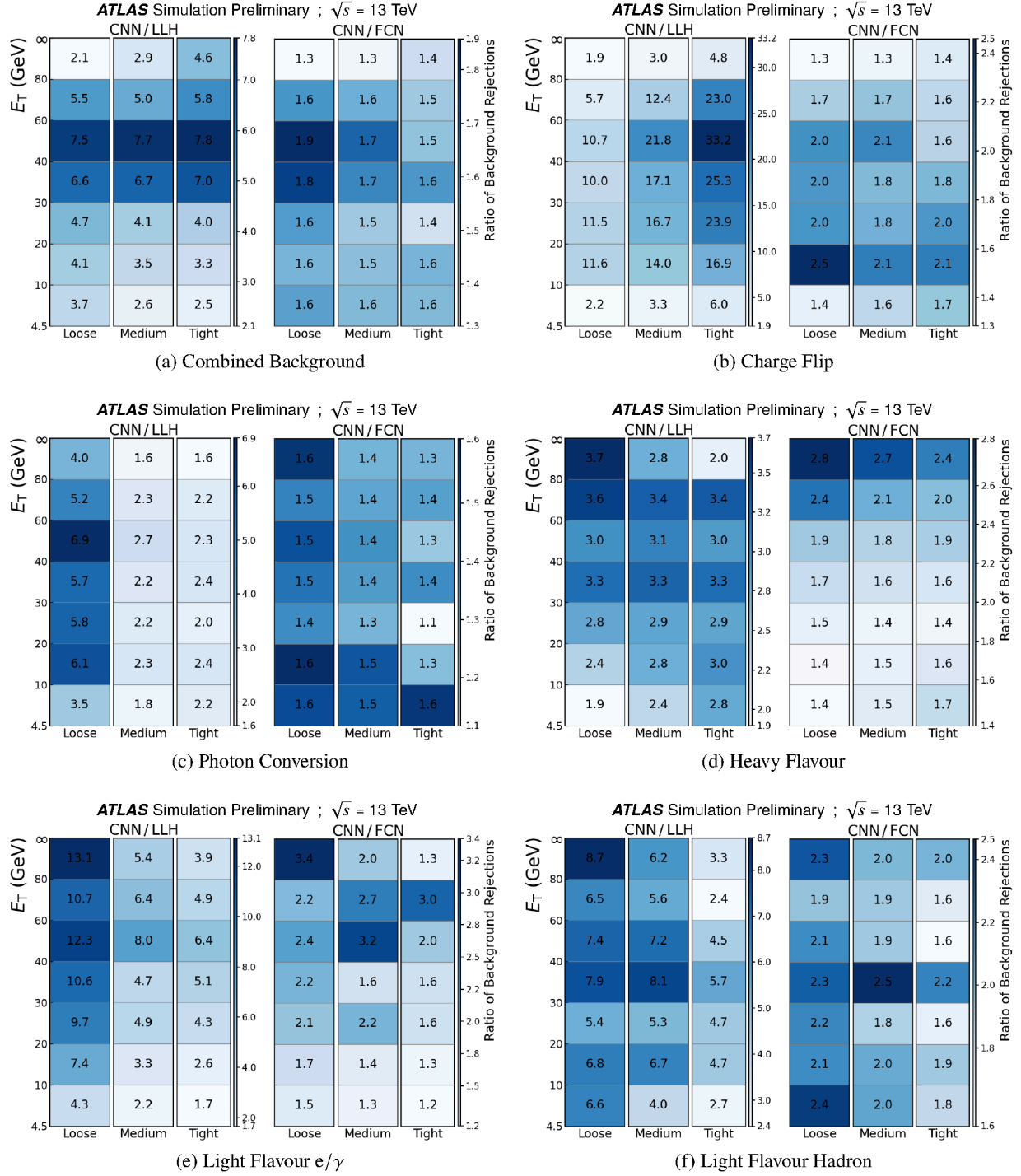


Figure 10: Ratios of background rejections obtained for the CNN with respect to the ones obtained for the LLH and FCN as a function of E_T (inclusively in η).

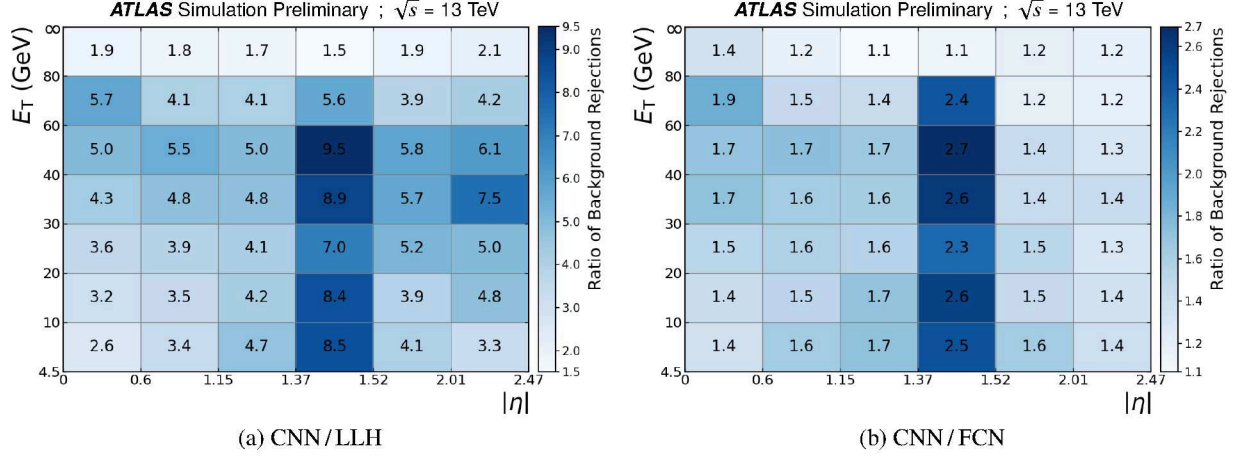


Figure 11: Ratios of background rejections obtained for the CNN (HLV+tracks+images) with respect to the ones obtained for (a) the LLH and (b) the FCN as a function of E_T and η for a signal efficiency corresponding to the LLH Loose working point.

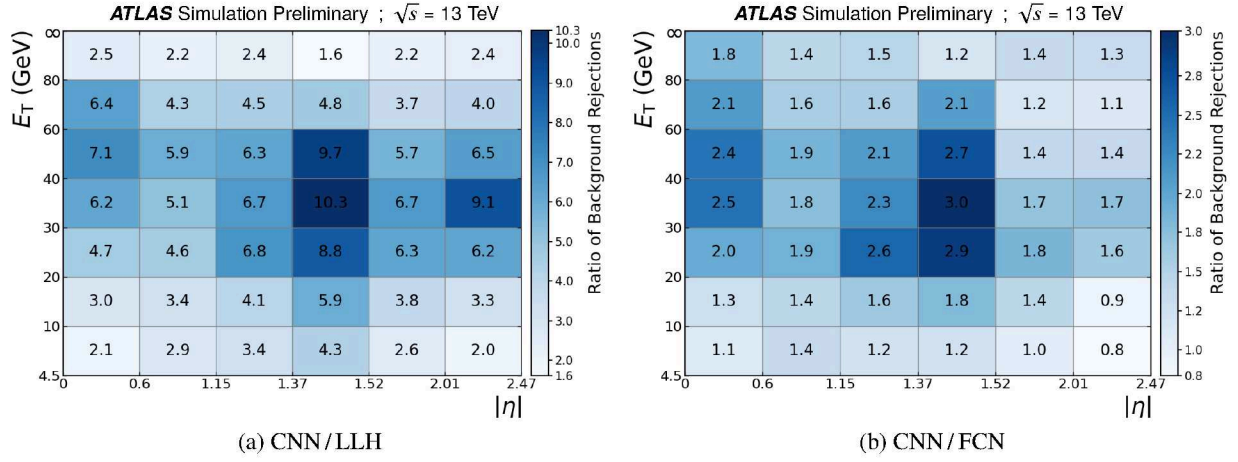
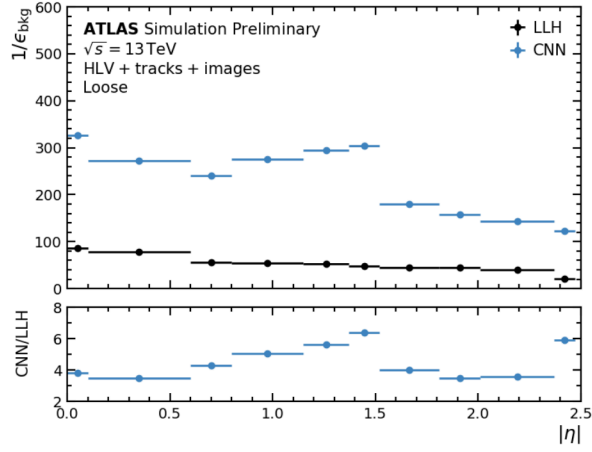
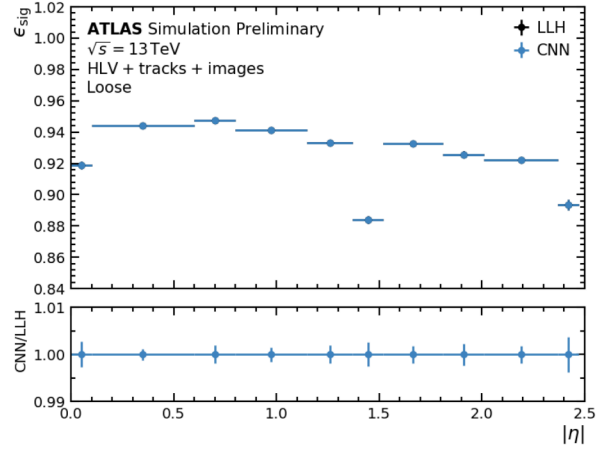


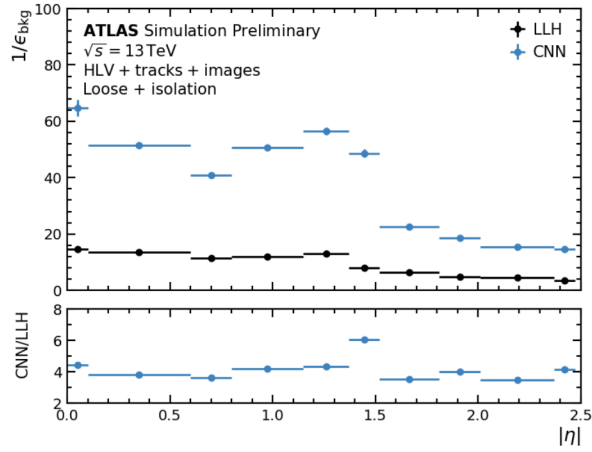
Figure 12: Ratios of background rejections obtained for the CNN (HLV+tracks+images) with respect to the ones obtained for (a) the LLH and (b) the FCN as a function of E_T and η for a signal efficiency corresponding to the LLH Loose working point. Isolation criteria (described in the text) have been applied to the electron candidates in addition to the CNN, LLH or FCN identification criteria.



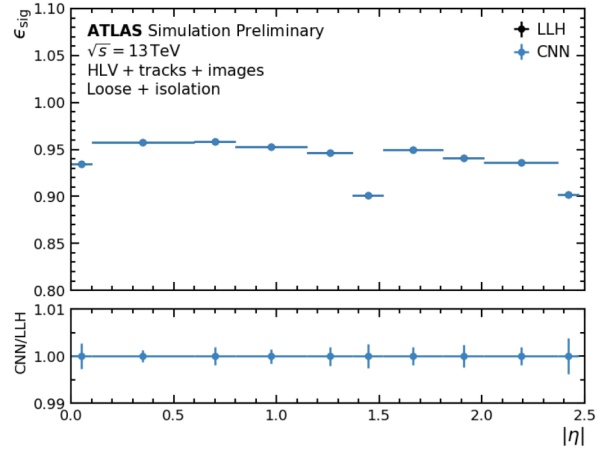
(a) Background Rejection (Loose)



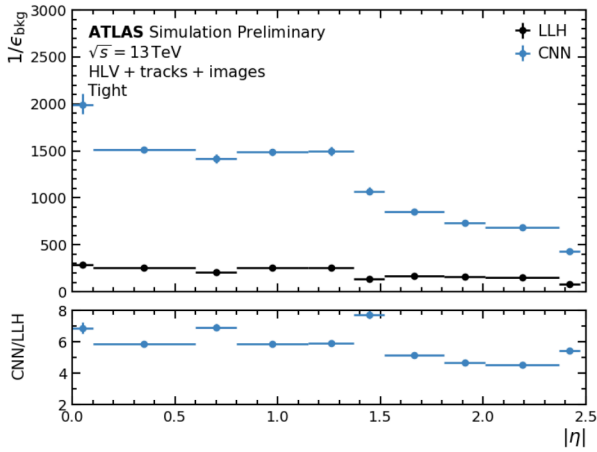
(b) Signal Efficiency (Loose)



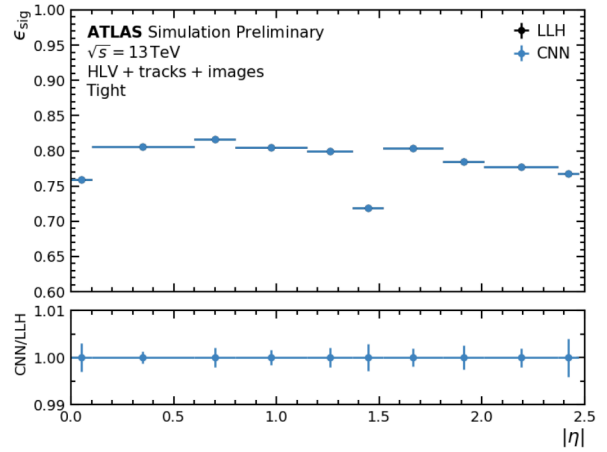
(c) Background Rejection (Loose + isolation)



(d) Signal Efficiency (Loose + isolation)

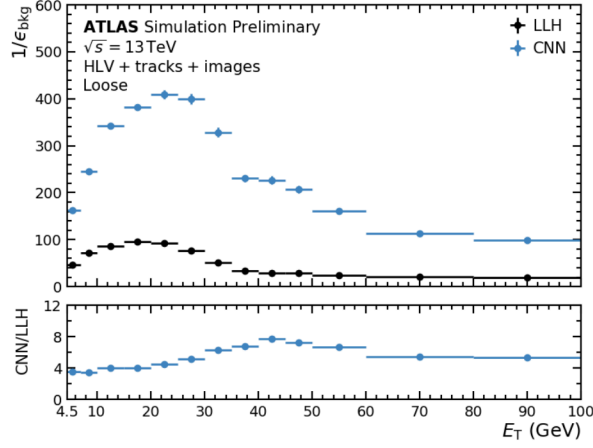


(e) Background Rejection (Tight)

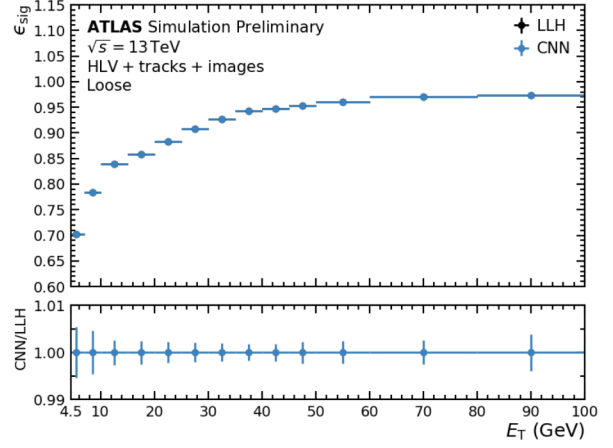


(f) Signal Efficiency (Tight)

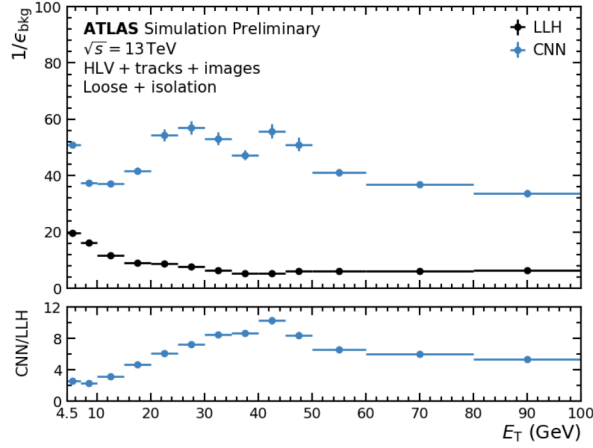
Figure 13: Background rejection and signal efficiency as a function of $|\eta|$ for the CNN and LLH while targeting the bin-by-bin LLH signal efficiencies of the (a)(b) Loose, (c)(d) Loose + isolation and (e)(f) Tight working points. In Figures (c)(d), the efficiency is computed with respect to isolated electrons. The lower panels show the ratio of these values for the CNN over the LLH (which is consistently equal to one for the signal efficiency of the CNN since it is adjusted to match the one of the LLH in each bin).



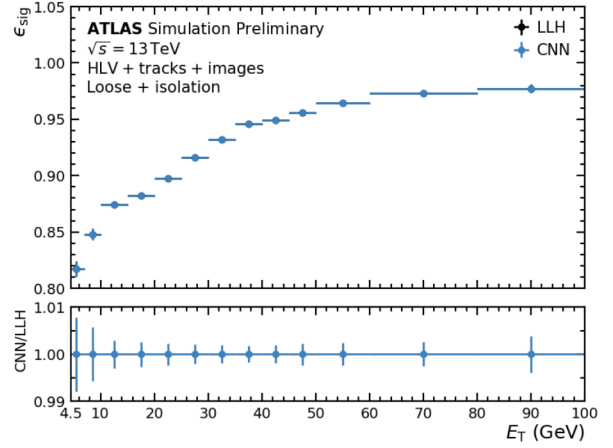
(a) Background Rejection (Loose)



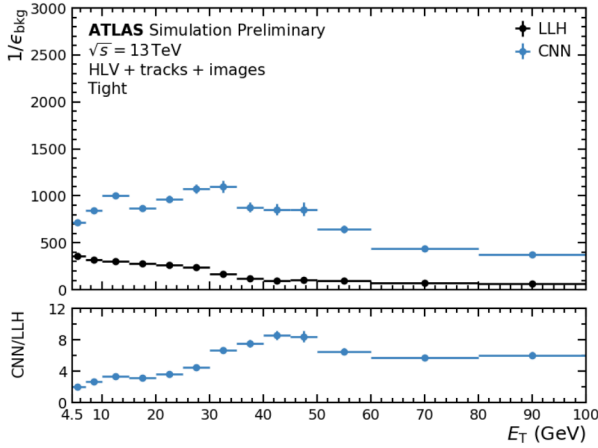
(b) Signal Efficiency (Loose)



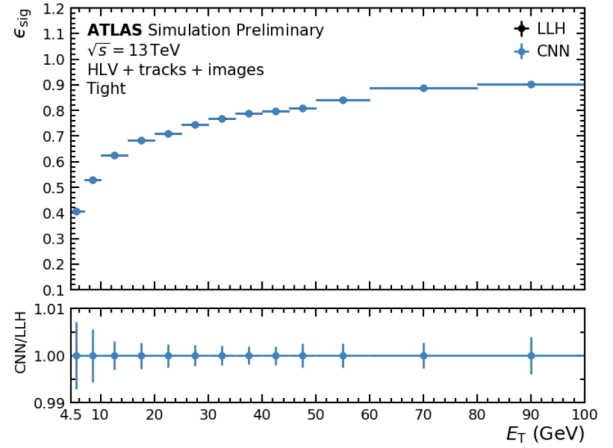
(c) Background Rejection (Loose + isolation)



(d) Signal Efficiency (Loose + isolation)

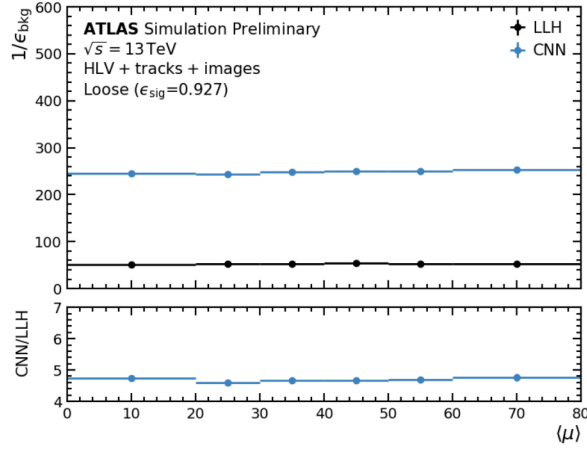


(e) Background Rejection (Tight)

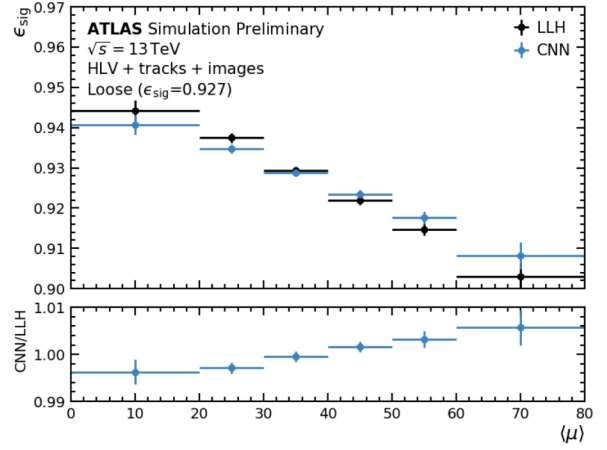


(f) Signal Efficiency (Tight)

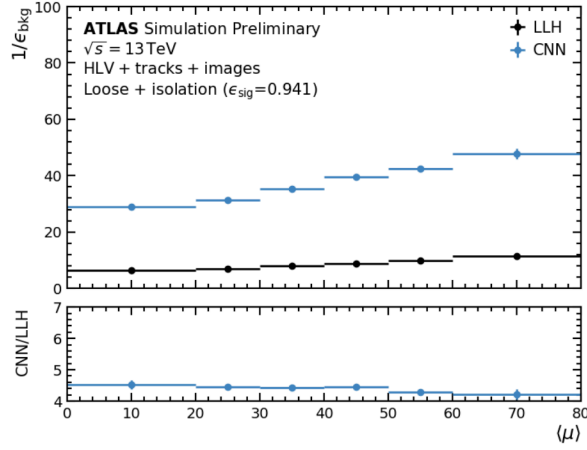
Figure 14: Background rejection and signal efficiency as a function of E_T for the CNN and LLH while targeting the bin-by-bin LLH signal efficiencies of the (a)(b) Loose, (c)(d) Loose + isolation and (e)(f) Tight working points. In Figures (c)(d), the efficiency is computed with respect to isolated electrons. The lower panels show the ratio of these values for the CNN over the LLH (which is consistently equal to one for the signal efficiency of the CNN since it is adjusted to match the one of the LLH in each bin).



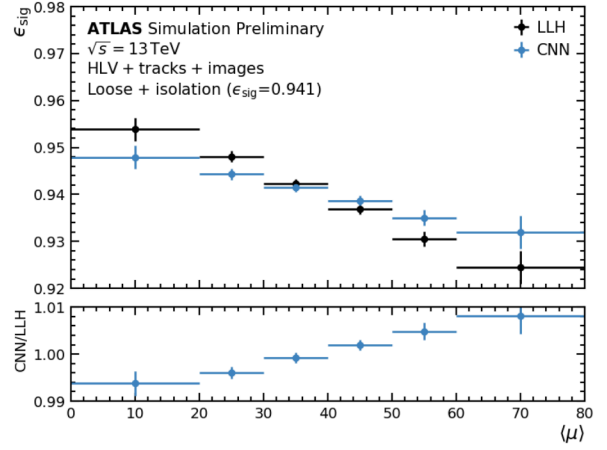
(a) Background Rejection (Loose)



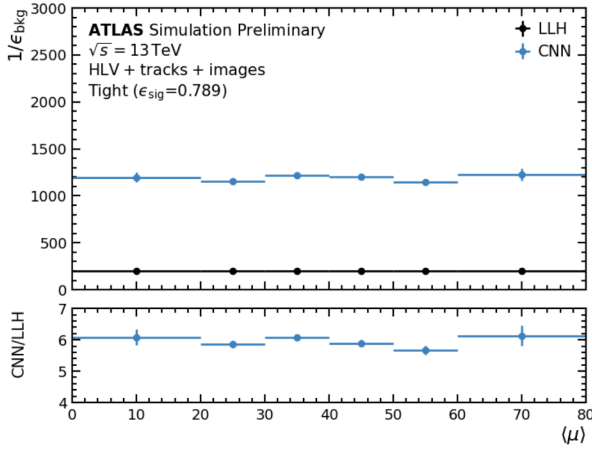
(b) Signal Efficiency (Loose)



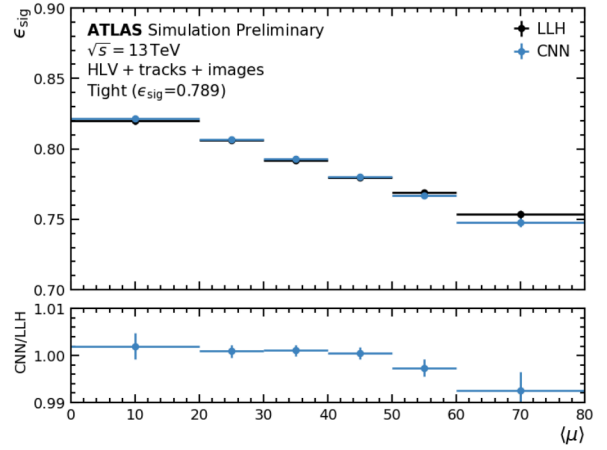
(c) Background Rejection (Loose + isolation)



(d) Signal Efficiency (Loose + isolation)



(e) Background Rejection (Tight)



(f) Signal Efficiency (Tight)

Figure 15: Background rejection and signal efficiency as a function of $\langle\mu\rangle$ for the CNN and LLH while targeting the integrated LLH signal efficiencies of the (a)(b) Loose, (c)(d) Loose + isolation and (e)(f) Tight working points. In Figures (c)(d), the efficiency is computed with respect to isolated electrons. The lower panels show the ratio of these values for the CNN over the LLH.

same samples. Although imperfect, this method is capable to provide a qualitatively faithful prediction of the electron class composition of a sample. Figure 16(b) shows that it can predict the true electron class ratios within about 5% for most of the class and the samples considered.

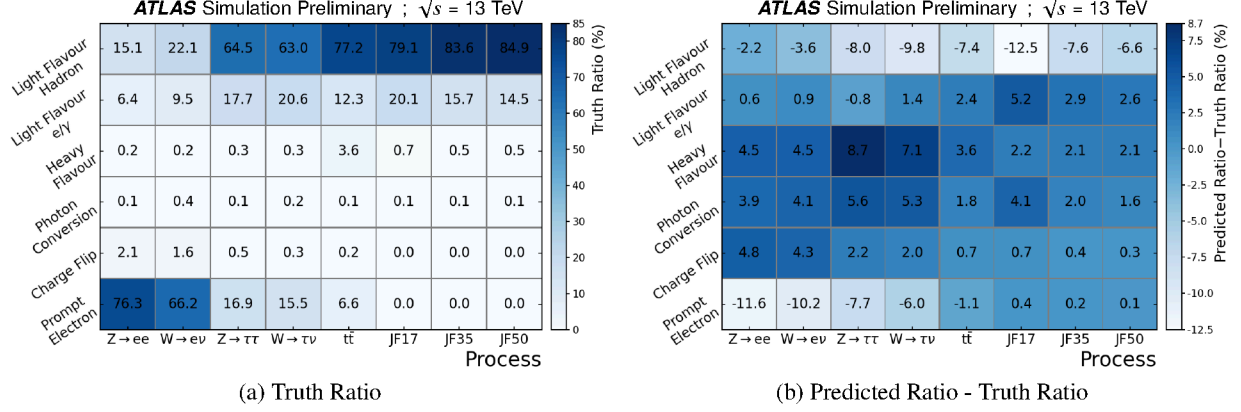


Figure 16: (a) True electron class fraction and (b) difference between the predicted and true class fraction as a function of the true electron class for a few representative simulated processes.

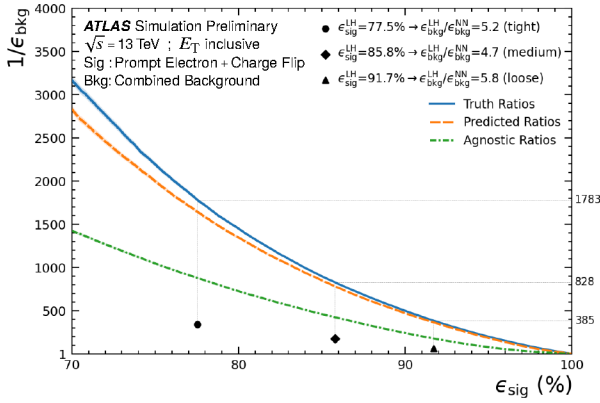
This feature of the CNN could be beneficial for physics analyses, such as guiding the choice of selection requirements of an analysis or to improve its background estimate. Furthermore, it could be used to enhance the identification performance of the CNN by predicting the weights to be used in Equation 1, which are found to be optimal when the weights correspond to the true electron class ratios. The CNN would be run once on a given sample to obtain the class ratios as explained previously. The discriminant could then be computed using the predicted ratios for the values of the weights that appear in Equation 1. The identification could then be performed on the sample by determining which electrons pass a selected threshold in the newly formed discriminant. This procedure requires to run the CNN only once to obtain the probability vector for each electron. Table 7 shows the improvement obtained in background rejection for a 90% signal efficiency for a few representative samples by using such a method compared to the case of an agnostic discriminant, i.e. when the weights of Equation 1 are all set to the same value. Significant improvement can be obtained, especially for samples that are enriched in electron backgrounds. That table also shows the additional improvement that could be obtained if one knew the true electron class ratio in a given sample. Further improvements could be achieved in theory but, for most samples, the majority of the achievable improvement is already obtained by computing the predicted electron class ratios using the CNN. This is also illustrated in Figure 17(a) (charge-flip electron included in signal) and 17(b) (charge-flip electron included in background) that show the ROC curves for signal efficiency versus the combined electron background when the agnostic, predicted and truth class ratios are employed when computing the CNN discriminant. They show that nearly optimal performance can be obtained when using the electron class ratios predicted by the CNN.

7 Conclusion

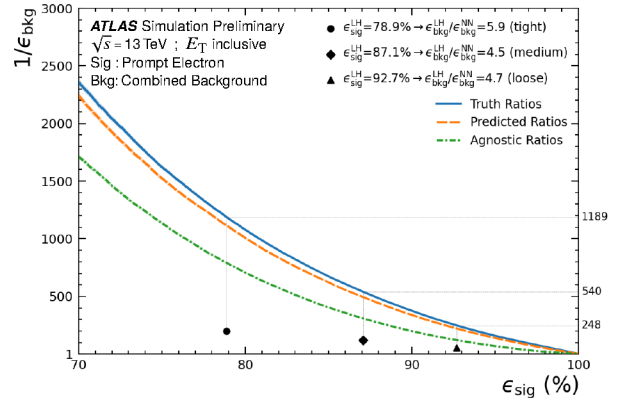
An electron identification deep learning algorithm based on convolutional neural networks has been developed in the context of the ATLAS experiment. In addition to using as input the same high-level variables that are used by the existing LLH and DNN algorithms, the CNN uses images of energy deposited

Table 7: Relative improvement in background rejection when using the predicted class ratio in the equation of the CNN discriminant with respect to using agnostic weights for a signal efficiency of 90%. The additional improvements that is in theory achievable if one could know the true class ratio is also shown.

Process	Agnostic \rightarrow Prediction	Prediction \rightarrow Truth
$Z \rightarrow ee$	9 %	21 %
$W \rightarrow e\nu$	26 %	27 %
$Z \rightarrow \tau\tau$	96 %	184 %
$W \rightarrow \tau\nu$	135 %	166 %
$t\bar{t}$ (at least one W boson $\rightarrow \ell\nu$)	67 %	6 %
JF17	172 %	58 %
JF35	290 %	44 %
JF50	313 %	80 %



(a) Combined Background (5-class)



(b) Combined Background (6-class)

Figure 17: ROC curves of the signal efficiency versus the combined electron background rejection obtained when agnostic (green), predicted (orange) and truth (blue) class ratios are employed when computing the CNN discriminant weights. The left and right figures show the case where charge-flip electrons are included in the signal and background, respectively.

in each of the electromagnetic and hadronic calorimeter layers, as well as the information of additional tracks matched to electron candidates. This ancillary information results in a large improvement in the electron identification performance. For example, for a signal efficiency corresponding to the Loose working point of the LLH, the combined background rejection for a fixed signal efficiency is improved by factors of approximately 2 to 10 with respect to the LLH algorithm depending on the electron kinematics. This is comparable or better than the background rejection obtained with the LLH Tight working point, but for signal efficiencies that are approximately 15-25% higher (in absolute term). Excellent signal versus background discrimination is particularly achievable against charged hadrons (e.g. π^\pm) faking the electron signature, for which background rejection factors approaching 10^4 are obtainable, about five times better than the LLH under the same conditions. For a signal efficiency corresponding to the Loose LLH working point, the information provided by the calorimeter images and additional tracks improves the background rejection by factors varying between 1.2 and 3.0 (depending on the electron kinematics and background class) with respect to a fully-connected network that uses as input only high-level variables. Furthermore, the CNN outputs a vector of probabilities that an electron candidate belong to each of six classes of signal

and background electrons that are defined. This feature can be used to determine the composition of an application sample, as tested in a variety of processes. In principle, this knowledge of an application sample composition can then be inserted in the computation of the final discriminant of the CNN, which significantly improves the performance over a final discriminant in which all of the electron classes would be weighed equally.

The results presented in this note have been obtained under the ideal conditions that the same simulated samples were used for both training the CNN and testing its performance. Since the CNN uses low-level detector information (such as calorimeter images), it might be particularly sensitive to detailed differences between real data and the simulation, such as noise level, dead cells or the modelling of particle interactions with matter. This challenge could be overcome by training the CNN using real data, which would necessitate to design control regions for each electron class. Another option would be to keep on using simulation as the main training sample, but provide some training examples coming from real data for which an adversarial network would be used to mitigate differences between real data and simulation [40].

References

- [1] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019) 639, arXiv: [1902.04655 \[hep-ex\]](#) (cit. on pp. 2, 8, 17).
- [2] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, *JINST* **14** (2019) P12006, arXiv: [1908.00005 \[hep-ex\]](#) (cit. on pp. 2, 8).
- [3] ATLAS Collaboration, *Identification of electrons using a deep neural network in the ATLAS experiment*, ATL-PHYS-PUB-2022-022, 2022, URL: <https://cds.cern.ch/record/2809283> (cit. on pp. 3, 15).
- [4] L. de Oliveira, B. Nachman, and M. Paganini, *Electromagnetic Showers Beyond Shower Shapes*, *Nucl. Instrum. Meth. A* **951** (2020) 162879, arXiv: [1806.05667 \[hep-ex\]](#) (cit. on p. 3).
- [5] Y. LeCun et al., “Handwritten Digit Recognition with a Back-Propagation Network”, *Advances in Neural Information Processing Systems*, ed. by D. Touretzky, vol. 2, Morgan-Kaufmann, 1989, URL: <https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf> (cit. on p. 3).
- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-based learning applied to document recognition*, *Proceedings of the IEEE* **86** (1998) 2278 (cit. on p. 3).
- [7] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003 (cit. on p. 3).
- [8] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633> (cit. on p. 3), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: <https://cds.cern.ch/record/1451888>.

- [9] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, *JINST* **13** (2018) T05008, arXiv: [1803.00844 \[physics.ins-det\]](#) (cit. on p. 3).
- [10] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* **77** (2017) 317, arXiv: [1611.09661 \[hep-ex\]](#) (cit. on p. 4).
- [11] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: <https://cds.cern.ch/record/2767187> (cit. on p. 4).
- [12] ATLAS Collaboration, *Monte Carlo Generators for the Production of a W or Z/ γ^* Boson in Association with Jets at ATLAS in Run 2*, ATL-PHYS-PUB-2016-003, 2016, URL: <https://cds.cern.ch/record/2120133> (cit. on p. 6).
- [13] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, arXiv: [hep-ph/0409146](#) (cit. on p. 6).
- [14] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, arXiv: [0709.2092 \[hep-ph\]](#) (cit. on p. 6).
- [15] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, arXiv: [1002.2581 \[hep-ph\]](#) (cit. on p. 6).
- [16] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO vector-boson production matched with shower in POWHEG*, *JHEP* **07** (2008) 060, arXiv: [0805.4802 \[hep-ph\]](#) (cit. on p. 6).
- [17] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, arXiv: [0710.3820 \[hep-ph\]](#) (cit. on p. 6, 7).
- [18] ATLAS Collaboration, *Measurement of the Z/ γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09** (2014) 145, arXiv: [1406.3660 \[hep-ex\]](#) (cit. on p. 6).
- [19] H.-L. Lai et al., *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024, arXiv: [1007.2241 \[hep-ph\]](#) (cit. on p. 6).
- [20] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, *JHEP* **07** (2002) 012, arXiv: [hep-ph/0201195](#) (cit. on p. 6).
- [21] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, *Eur. Phys. J. C* **45** (2006) 97, arXiv: [hep-ph/0506026](#) (cit. on p. 6).
- [22] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS Interface in C++: Technical and physics documentation*, *Comput. Phys. Commun.* **199** (2016) 86, arXiv: [1011.0937 \[hep-ph\]](#) (cit. on p. 6).
- [23] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152 (cit. on p. 6).
- [24] ATLAS Collaboration, *Simulation of top-quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2016-004, 2016, URL: <https://cds.cern.ch/record/2120417> (cit. on p. 6).

- [25] R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, arXiv: [1410.8849 \[hep-ph\]](#) (cit. on p. 6).
- [26] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: <https://cds.cern.ch/record/2216168> (cit. on p. 6).
- [27] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159, arXiv: [1410.3012 \[hep-ph\]](#) (cit. on p. 6).
- [28] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419> (cit. on p. 6).
- [29] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, arXiv: [1207.1303 \[hep-ph\]](#) (cit. on pp. 6, 7).
- [30] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: <https://cds.cern.ch/record/2206965> (cit. on p. 7).
- [31] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, URL: <https://cds.cern.ch/record/1099735> (cit. on p. 8).
- [32] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) 490, arXiv: [1603.02934 \[hep-ex\]](#) (cit. on p. 8).
- [33] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012, URL: <https://cds.cern.ch/record/1449796> (cit. on p. 8).
- [34] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825, URL: <https://arxiv.org/abs/1201.0490> (cit. on p. 11).
- [35] Martín Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015, URL: <https://www.tensorflow.org/> (cit. on p. 15).
- [36] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, ed. by Y. Bengio and Y. LeCun, 2015, URL: <http://arxiv.org/abs/1412.6980> (cit. on p. 15).
- [37] C. Cortes, M. Mohri, and A. Rostamizadeh, *L2 Regularization for Learning Kernels*, 2012, URL: <https://arxiv.org/abs/1205.2653> (cit. on p. 15).
- [38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, *Journal of Machine Learning Research* **15** (2014) 1929, URL: <http://jmlr.org/papers/v15/srivastava14a.html> (cit. on p. 15).
- [39] J. Neyman and E. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, *Philosophical Transactions of the Royal Society A* (1933) 289 (cit. on p. 15).

- [40] G. Louppe, M. Kagan, and K. Cranmer, “Learning to Pivot with Adversarial Networks”, *Advances in Neural Information Processing Systems*, ed. by I. Guyon et al., vol. 30, Curran Associates, Inc., 2017, URL: <https://proceedings.neurips.cc/paper/2017/file/48ab2f9b45957ab574cf005eb8a76760-Paper.pdf> (cit. on p. 30).