LIP-STUDENTS-20-17

Search for dark matter and supersymmetry using machine learning at SHiP

Francisco Safara[1,a] and Raúl Santos[2,b]

[1] *Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal*
[2] *Instituto Superior Técnico, Lisboa, Portugal*

Project supervisors: N. Leonardo, G. Soares

October 2020

Abstract. SHiP is an Intensity Frontier experiment aimed at the search for particles with extremely feeble interactions, low masses and long lifespan. Such particles are predicted in a number of recently elaborated scenarios of the *hidden sector* of particle physics. In this project we used the SHiP software framework to simulate hidden particles, specifically dark photons and neutralinos, and study their kinematic properties. We have implemented and tested several machine learning techniques, with the aim of rejecting the neutrino background while maintaining a high signal efficiency. We were able to achieve, exploring neural networks with feature pre-processing, regression and classification, nil backgrounds and signal efficiencies above 95%.

KEYWORDS: Hidden Sector, Dark Photons, Neutralinos, Neural Networks

1 Introduction

The Standard Model (SM) of particle physics aims to describe the most fundamental properties of matter, it was developed during the second half of the 20th century, in a global initiative based on the ideas of unification and symmetries. It has provided a consistent description of Nature's fundamental constituents and interactions.

However, it fails to explain a number of observed phenomena in particle physics, astrophysics and cosmology such as the matter–antimatter asymmetry, the nature of dark matter and dark energy. To explain this phenomena, newer particles and/or interactions would be needed, but until this moment no direct experimental evidence exists.

One possible reason for why these hypothetical particles have not yet been observed is that they are too heavy and require higher collision energies to be detected. This research is pursued through the so-called *energy frontier*, namely at CERN's LHC.

Another possibility is that their interactions with SM particles are extremely feeble. Some examples of such particles that have been theorized in recent years include: Heavy Neutral Leptons (HNL), Dark Photons (DP), and Neutralinos. Here, different kinds of experiments are needed, that explore the *intensity frontier*. This is the case of the SHiP experiment [1], that is under development and provides the framework for the research work here presented. The overarching goal is to explore the landscape of new physics possibilities, as in Fig. 1, through complementary experimental avenues.

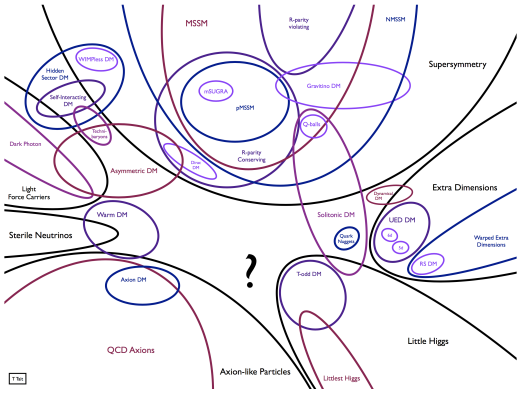

Figure 1: Theory landscape of dark matter candidates [2].

2 The SHiP experiment

SHiP (Search for Hidden Particles) is a new general-purpose fixed-target experiment designed to use the high-intensity beam of protons available from the CERN Super Proton Synchrotron (SPS) accelerator. The target goal is to search for very weakly interacting, long-lived particles, with masses in the range from hundreds of MeV to few GeV, inaccessible in other experiments, as well as to make measurements involving the tau neutrino. Hidden particles are predicted by a large number of models beyond the Standard Model, and include Heavy Neutral Leptons (HNL), Dark Photons (DP), and the Neutralinos [3].

A schematic of the SHiP detector is shown in Fig. 2. It consists of a target, where a beam of high intensity (400 GeV) protons from the SPS accelerator hit, followed by a hadron absorber. The muons produced in the collision are mostly deflected by the magnetic field generated in the muon shield. Then there is a scattering and neutrino detector, followed by a muon identification system and a timing detector. The vacuum decay vessel, with a length of 50m and a low pressure of 1 mbar, is where the hidden sector particles, generated from the proton collisions with the target, are expected to travel to and decay on. The decay products are identified and their kinematic properties quantified upstream, on the straw trackers, timing detector, electromagnetic calorimeter, and muon identification system.

A prototype of the timing detector has been developed at LIP with the resistive plate chamber (RPC) technology. High efficiencies and precise timing measurements, of or-

[a] e-mail: franciscosafara@outlook.com
[b] e-mail: raulfilipe1998@hotmail.com

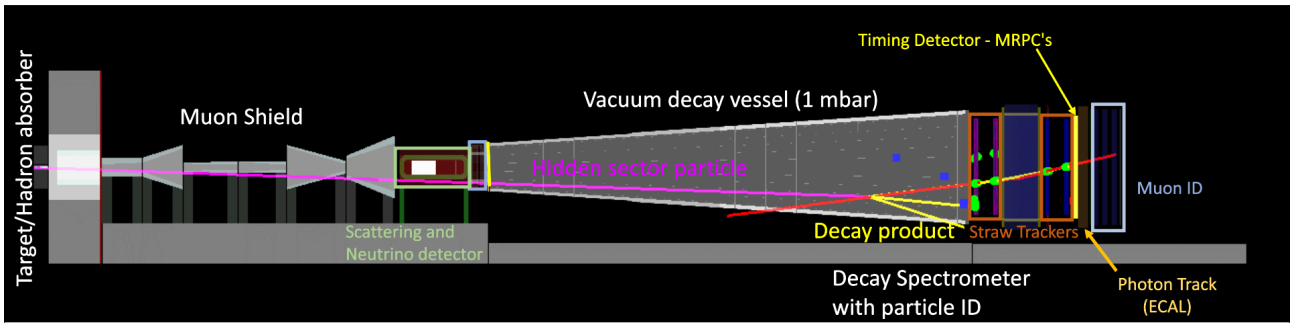

Figure 2: The SHiP experiment. The event display illustrates the decay of a simulated hidden-sector particle, and the reconstruction of its decay products from the signals in the sub-detector elements. The Z coordinate is horizontal along the page, the Y is vertical to the page, X is away from the plane of the page.

der 50ps, are achieved. Figure 3 displays the experimental setup used for measuring the performance in beam tests at CERN.

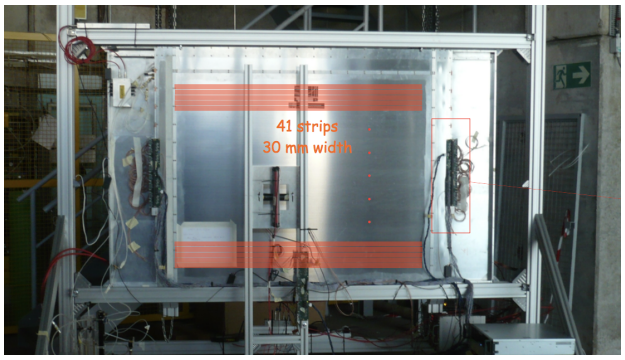

Figure 3: The SHiP RPC detector prototype developed at LIP, being exposed to negative pions of 8 GeV in a beam test at CERN [4].

3 Simulation of Hidden Sector particles

Computational simulation is necessary to study the behavior of particles and their decays, and propagation through the detector elements, in order to optimize the design and construction of the experiment, and estimate its sensitivity to the various hidden-sector signals. With this goal in mind, a Monte Carlo simulation and software framework was created for the SHiP experiment, FairShip, which was developed based on the FairRoot software environment. In FairShip simulations, primary collisions of protons are generated with Pythia8, and the subsequent propagation and interactions of particles are simulated with GEANT4. Neutrino interactions in the decay vessel are simulated with GENIE, and heavy flavor production and inelastic muon interactions with Pythia 6 and GEANT4 [5].

3.1 Dark Photon

The Dark Photon (DP) is a new gauge boson whose existence is only theoretical, that, if exists, would be an force carrier, like the photon in electromagnetism, represented by A'. But instead of working between charged particles of regular matter, like a photon does, the dark photon would enable interactions between particles of dark matter.

The dark photon could be detected because of its kinetic mixing with the ordinary, visible photon. Three different mechanisms are possible for the production of such new particles at a fixed target experiment: via the bremsstrahlung process, via Drell-Yan like QCD processes, or through meson decay channels involving photons. After having been produced, the DP may decay into hadrons, e^+e^- (see Fig. 4), $\mu^+\mu^-$ and $\tau^+\tau^-$[6].

The simplest DP theoretical model is characterized by two parameters: the mass of the dark photon $m_{A'}$ and its kinetic mixing parameter with the photon, ϵ. These two quantities parameterise therefore the relevant phase space, $(\epsilon^2, m_{A'})$, shown in Fig. 5.

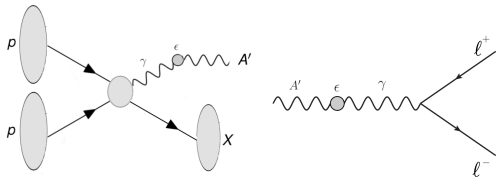

Figure 4: Feynman diagrams of dark photon production (left) and decay (right) through kinetic mixing.

3.2 Neutralinos

Supersymmetry (SUSY) is a popular SM extension, that connects bosons with fermions, by giving each boson a fermionic (super)partner and vice versa. It is a well motivated scenario, alleviating several challenges to SM, e.g. hierarchy problem and grand unification. Experimentally, however, the more standard (and natural) SUSY scenarios are under pressure, motivating the exploration of alternative signatures. In favoured SUSY scenarios, the neutralino is the lightest SUSY particle (LSP), a stable, dark matter candidate. The mass range 0.7 eV to 24 GeV is however excluded for stable LSP, motivating its explo-

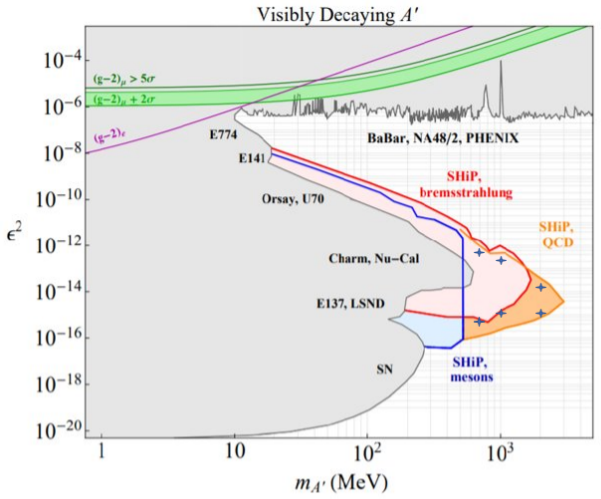

Figure 5: The dark photon (ϵ^2,$m_{A'}$) phase space. The gray area is excluded by previous experiments, and the predicted SHiP sensitivity is shown, for each production mode [1]. The (blue) crosses are the points chosen for the simulations, using the FairShip software.

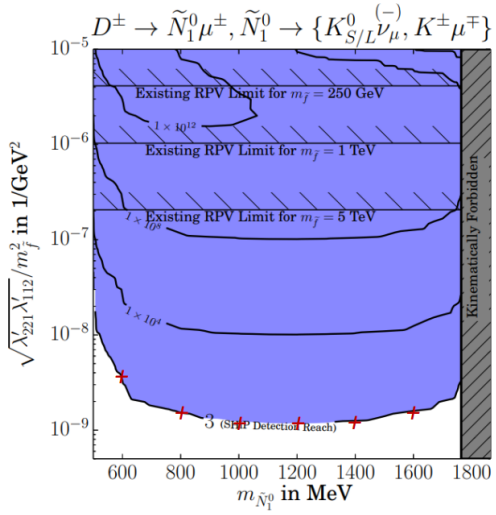

Figure 7: The neutralino mass-coupling phase space, with the SHiP expected sensitivity, for scenarios with the indicated sfermion masses [7]. The points chosen for our simulations are indicated by the (red) crosses.

ration through R-parity violation signatures. This is what will be done in this work.

The neutralino is produced from charm meson decays, represented in the Feynman diagrams in Fig. 6. We will consider the following neutralino decays and attempt their reconstruction and analysis:

$$\tilde{N}_1^0 \to K_S^0 \nu_\mu \quad \text{and} \quad \tilde{N}_1^0 \to K^\pm \mu^\mp .$$

In order to generate the particle we needed to specify the mass, the coupling constant, and the sfermion mass. In Fig. 7, the values of masses and couplings employed in our simulations are shown. The sfermion mass was fixed at 1 TeV. The points were picked according to SHiP experiment sensibility range - the coordinates in the phase space that correspond to a detection rate of 3 events over 5 years.

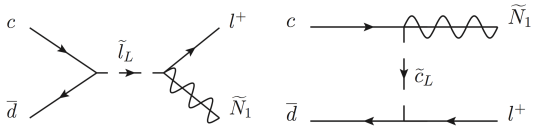

Figure 6: Feynman diagrams for neutralino production from charm mesons, $D^+ \to \tilde{N}_1^0 \ell^+$.

3.3 Neutrino DIS background

Neutrinos that originate from the target may propagate through to the decay vessel, and interact with the air of its near vacuum. This interaction is the deep inelastic scattering, and it happens when, in this case neutrinos (but any lepton in general), scatter off hadrons, shattering them and emitting many new particles in the process. These new particles can then be mistaken as decay products of hidden sector particles, giving off a potentially false signal towards the detection of these new and rare particles.

We simulated 4 million total neutrino DIS events, with 1 million DIS events of electron neutrino (ν_e), muon neutrino (ν_μ) and corresponding anti-neutrinos.

4 Observables and baseline selection

From the simulations, for each event, we can recover many of its kinematic properties (like momentum of the mother particle, momentum of daughter particles, opening angle, ...). Following up we explain the extracted properties that we used for selection. The momentum properties are applied to the mother particle, that decays to the daughter particles.

- **Total momentum**
- **Transverse momentum**: the component of the momentum of the new particle that is orthogonal to the direction of travel of the originating particle (daughter: relative to the mother particle; mother: relative to the incident beam).
- **Fraction of transverse momentum**: ratio between transverse momentum and total momentum.
- **Opening angle**: angle between the daughters momenta (we'll always have only 2 daughter particles).
- **Impact parameter**: Propagating the mother particle back onto the target, it is the distance between where on the target the particle originated and where the proton beam struck.
- **Decay vertex position** (X, Y, Z): Space coordinate where the mother particle decayed.

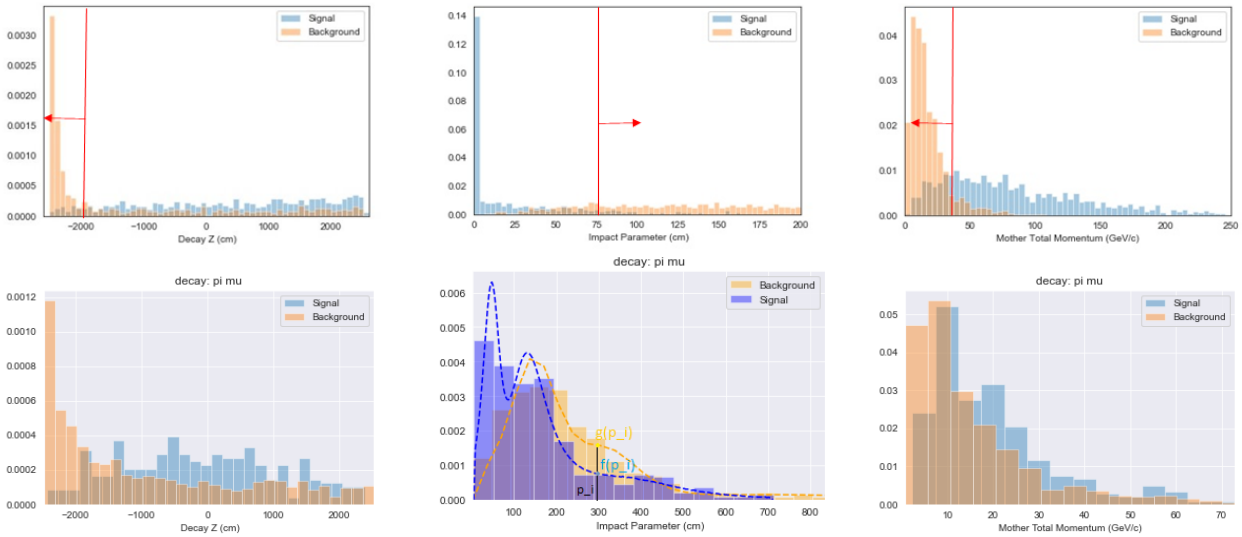

Figure 8: Distributions of selected kinematic properties for dark photons (top) and neutralinos (bottom) signals (blue) versus the neutrino background (orange). The distributions are normalized to a common area.

4.1 Dark Photons distributions

For each of the 6 phase space points shown in Fig. 5 we simulated 4000 events. These points were chosen in the limit of the expected sensitivity, just accessible with the SHiP experiment. In the simulations, the production and decay modes need to be specified. We focused on QCD production (masses between 0.8 and 2.0 MeV), and the hadronic decay modes leading to charged particles.

After the simulation, the DP decays, that were reconstructed, were separated according to the hadronic final states, and their observed branching fractions assessed. We can observe, in Table 1, a clear abundance of $\pi^{\pm}\pi^{\mp}$ decays, totaling about 44.6% of the total reconstructed particles (about 7194 events). These were the ones that were further analyzed. A number of 1099 background events were also simulated with $\pi^{\pm}\pi^{\mp}$ decays.

Table 1: Observed branching fractions of the reconstructed dark photon decays.

decay	frequency
$\pi^{\pm}\pi^{\mp}$	44.6%
$X\pi^{\pm}\pi^{\mp}$	20.6%
$K^{\pm}\pi^{\mp}$	4.4%
Other	30.4%

4.2 Neutralino distributions

For each of the masses shown in Fig 7 and for each of the considered decays ($\tilde{N}_1^0 \to K_S^0\nu_\mu$ or $\tilde{N}_1^0 \to K^{\pm}\mu^{\mp}$), 3000 events were simulated. From these 36000 events, only 7366 events were usable (many events fall outside of the detector acceptance, and get discarded, as they can't be reconstructed). Table 2 shows the branching fractions for the reconstructed events.

Following the event reconstruction, the four main reconstructed final states were further analyzed: $\pi^{\pm}\pi^{\mp}$, $\pi^{\pm}\mu^{\mp}$, $K^{\pm}\mu^{\mp}$ and $\mu^{\pm}\mu^{\mp}$. These form our signal. The background will be the neutrino deep inelastic scattering events that result in the same reconstructed final state particles.

Table 2: Observed branching fractions of the reconstructed neutralino decays.

$K_S^0\nu_\mu$	
decay	**frequency**
$\pi^{\pm}\pi^{\mp}$	89%
$\pi^{\pm}\mu^{\mp}$	9.2%
$\mu^{\pm}\mu^{\mp}$	0.5%
other	1.4%

$K^{\pm}\mu^{\mp}$	
decay	**frequency**
$K^{\pm}\mu^{\mp}$	84%
$\mu^{\pm}\mu^{\mp}$	9.1%
$\pi^{\pm}\mu^{\mp}$	2.7%
$\pi^{\pm}\pi^{\mp}$	1.9%
other	2.5%

4.3 Cut-based selection

The conventional, most natural and simpler way to separate signal from background is to apply a set of independent cuts on the individual variables. We find among the physical quantities of each event those that are more *discriminating* and we apply cuts on these variables or on combinations of these variables.

The distributions of a selected set of variables formed from the DP data are shown in Fig. 8 (top row). The vertical (red) lines display the cut value, and the arrow indicates the discarded range. These cuts are usually arrived at by a process of trial and error, informed by common sense and physics insight.

Table 3: Cut-based selection results for dark photon particles that decay into $\pi^{\pm}\pi^{\mp}$. The Signal efficiency and number of Background surviving events after each cut are shown in the second and third columns, respectively.

Variable	Cut	Signal efficiency	Background yield
		1	1099
Impact parameter	<75 cm	0.81	157
Mother total momentum	>40 GeV	0.71	31
Decay vertex Z	>-2000 cm	0.66	1
Daughters total momentum	>3.5 GeV	0.64	0

The results obtained are summarized in Table 3. We can observe that, although a simple and fast method, the cut-based selection procedure can produce a satisfactory signal efficiency, typically larger than 60%.

A cut-based procedure is not in general the best option and there is no guarantee that the procedure will lead to the most optimal selection [8]. As our goal is to obtain signal efficiency values close to unit for zero background, "satisfactory" results are not good enough, and more sensitive and robust methods are therefore explored next.

5 Machine Learning

The cut-based selection method is a rather coarse approach, where a lot of signal may be lost on the absolute cuts that are applied. The logical step from there is to use a more adaptable approach, that has in consideration the correlation between the particle kinematic properties, is protected against human error and can take in consideration all of the feature complexity of data and properly select the signal. We chose to use a Binary Classification Neural Network [9].

For the implementation of our algorithms we used Keras, an open-source library that provides a Python interface for artificial neural networks. Keras acts as an interface for the TensorFlow library.

5.1 Classification Neural Network

Neural Networks (NN) have become a popular multivariate method employed in data analysis because of their power and ease of use. They are now used quite extensively in particle physics (and beyond).

A typical NN consists of an interconnected group of nodes arranged in layers, as in Fig. 9, where each node processes information and then passes the results to the next layer of nodes. The first layer, also know as input layer, receives the data features, and the last layer outputs the NN response. The connections between nodes are defined by a weight, whose values are adapted during the training phase. Neural networks learn by processing examples (training phase), each of which contains a known "input" and "result" forming probability-weighted associations between the two; the NN adjusts its weighted associations according to a Loss Function (e.g. mean square error between the actual output and the desired output), by

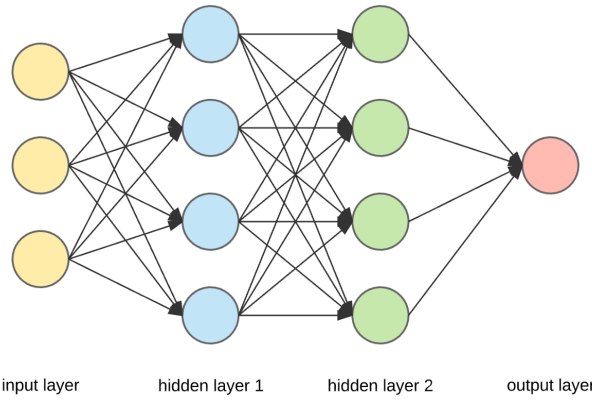

Figure 9: Typical structure of a Neural Network, where each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another.

trying to minimize the Loss value. Each time the algorithm sees the entire data set is called an epoch, the amount that the weights are updated during each epoch is referred to as the “learning rate”. A learning rate that is too large can cause the model to converge too quickly to a sub-optimal solution, whereas a learning rate that is too small can cause the process to get stuck.

The performance of a NN with just 2 outputs (Binary Neural Network) can be analyzed by the ROC curve. The ROC curve is the graph of the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings, as in Fig. 10. The larger the area under the ROC curve is, the most performing (in general) the NN is.

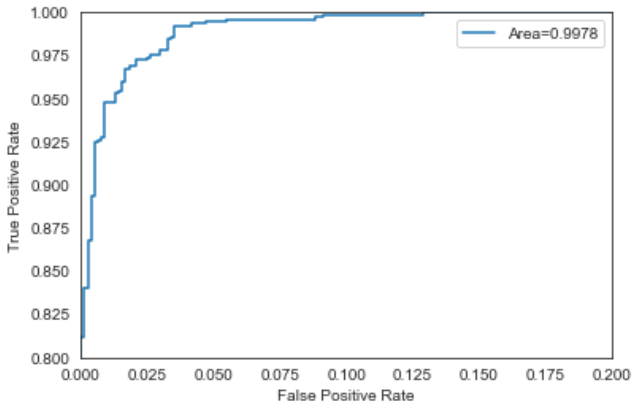

Figure 10: Example of ROC curve obtained after NN training in the DP analysis with an area under the curve of 0.9978.

We used a collection of the more relevant kinematic properties, including the Impact Parameter, Mother Total Momentum, Daughters Total Momentum, etc. as input data and we trained the NN model by providing it with labeled data: Signal (score=1) and Background (score=0). We divide the data into training and test sets, with an 80-20 split where the test set is used to check the performance.

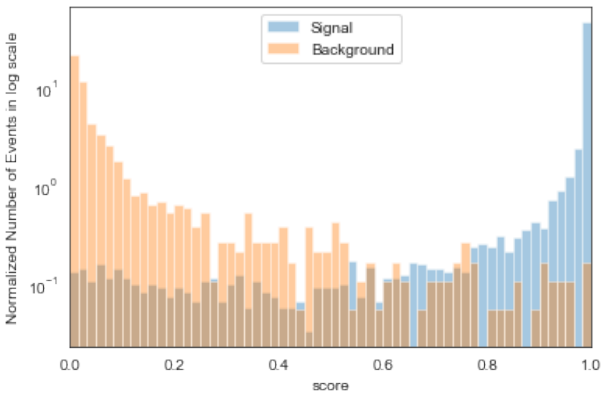

Figure 11: The normalized distribution (log scale) for the NN output score, applied to DP (signal, in blue) and DIS (background, in orange).

We used an 9 layers sequential model with a total of 300 epochs and a learning rate of 0.001.

5.2 Dark photon: selection with standard binary classification

To avoid imbalanced data sets, as the amount of simulated background and signal differ, we decided to divide the data into a symmetric subset (under sampling), where the extra data is employed for further validation. (For example, if we had 1000 signal and 1500 background, we would use 1000 signal and background for training and validation, and the extra 500 background events would only be considered as extra validation for the NN metrics.)

Applying the simple binary NN to the analysis of the DP (signal) and DIS (background) data, Fig. 11 shows that a good degree of separation is achieved. That is, the NN model can predict with accuracy if the Mother particle is Signal or Background. Desired levels of signal purity and background rate may be attained for different NN scores, as seen in the ROC curve in Fig. 10.

In Fig. 12 we can observe the behavior of the Loss function. As expected, the Loss decreases after each epoch, for the training data, but, for the testing data, after some epochs, the Loss remains constant. That shows that our model does not improve by arbitrarily increasing the number of epochs.

We can ensure 0 background (no False Positives) by defining a decision threshold as the bigger score attributed to a background event, values with a score higher are guaranteed to be signal. Unfortunately, to have 0 background, a considerable number of events will got mistaken as background, this is the result of a major flaw of this method: although the majority of the data is labeled correctly, some results are completely off, evaluating a background event with a really high score (nearly 1), that leads to a considerable amount of discarded signal by the threshold.

The threshold, for the data shown in Fig. 11, was 0.9905, the final signal efficiency was 0.82 (18% of the total number of signal particles that were analyzed were discarded by the application of the threshold). It is important to notice that, if we were to accept a small number of background as False Positives (less than 1% of the background events), the threshold would take a considerably lower value, giving a signal efficiency of more than 0.90, what shows that the number of background events that were miss identified is small .

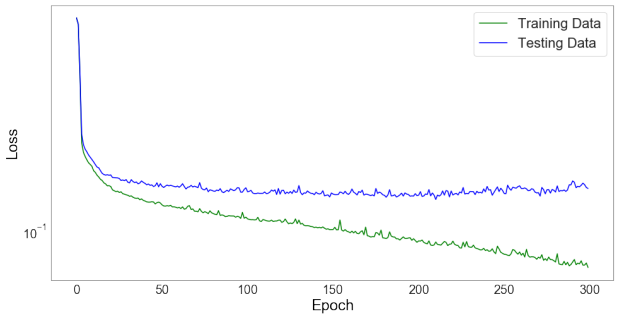

Figure 12: Loss function, obtained in the DP NN analysis of the Training Data (green) and of the Testing Data (blue) for a total of 300 epochs.

5.3 Feature pre-processing

Although we had good results (as we will see next) they were not the best, in hopes to find room for improvement, we tried another method which takes advantage of the approximated distribution function of the data and inputs the calculated probabilities onto a classification neural network, thinking that this transformation of the data may make the problem easier to the classifier.

For each kinematic property, we fitted a neural network onto the cumulative distribution function of the data. By using the computed losses and by comparing the finite difference to the data histogram, we were able to achieve, for the most part, satisfactory approximations to the distributions.[1]

For a certain kinematic property, and for each data point in it, we estimated the probability of being signal or background based on the fitted distribution. For the event i and for the associated kinematic property p_i, the probability of being signal is given by

$$P_i^{signal} = \frac{f(p_i)}{f(p_i) + g(p_i)} \,. \tag{1}$$

On Fig. 8 (bottom, center), it is shown the fitted impact parameter distribution, along with possible values for p_i, $f(p_i)$ and $g(p_i)$.

The functions $f(p)$ and $g(p)$ correspond to the fitted probability distributions of, respectively, the signal and the background. This probability analysis was done for every kinematic property, and so we obtained these *violin plots* (Fig. 13) with the probability of a given data point being

[1] Here a NN is used for *regressing* rather than *classifying* the data. More refined methods could be pursued to approximate the data distribution, such as kernel density estimation (KDE). Still, very good results are reached overall.

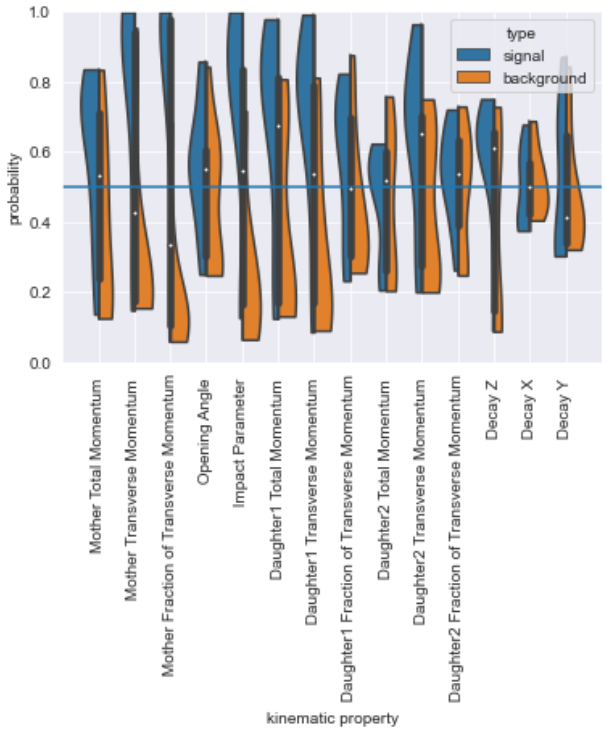

Figure 13: Violin plots comparing the estimated probabilities (P_i^{type}) of the many kinematic properties, for the neutralino $\pi^{\pm}\mu^{\mp}$ reconstructed decay. The horizontal (blue) line corresponds to a probability of 0.5. Isolated, these probabilities do not fully separate signal from background.

a signal, for signal events and for background events. Ideally, for all the kinematic properties and for all the background events, the probability would be below 0.5. This isn't always the case. Figure 13 shows a violin plot with an example of this transformation on the signal and background events. There are plenty of background events that, isolating the kinematic properties, could be taken as a signal.

As the distribution estimation is not ideal (due, for example, to the method or to the limited statistics of considered events), we take this probability data (the P_i^{type} where i goes through all simulated events) further into a binary classification neural network, in hopes that it can recognize useful patterns on these estimated probabilities.

5.4 Neutralinos: selection with feature pre-processing

For the neutralino we tried the method described in section 5.3. An example of a regression to the distribution function is shown in Fig. 8 (bottom, center). It was done to the Impact parameter of the $\pi^{\pm}\mu^{\mp}$ reconstructed decay. The loss function for the classification problem of $\pi^{\pm}\mu^{\mp}$ decay is shown in Fig. 14. Comparing the loss of the Training Data with the loss of the Testing Data it is clear that there is overfitting. This could be due to the complexity of the model (shape of the neural network - more nodes and layers → more complexity). To get better results, one has to either simplify the model or to just get more data.

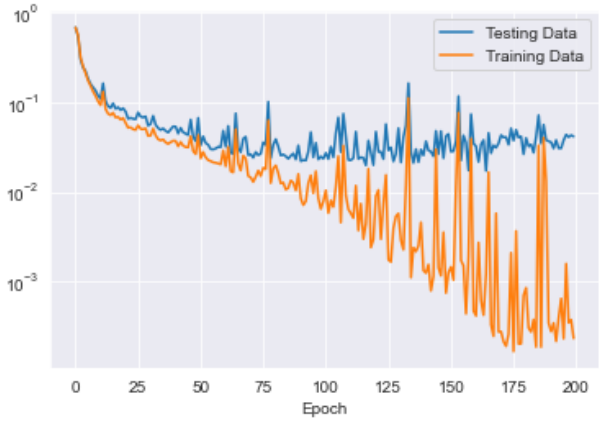

Figure 14: Loss function for the binary classification of the $\pi^{\pm}\mu^{\mp}$ decay (with feature pre-processing). The loss of the Training Data is in orange and of the Testing Data is in blue

The histogram in Fig. 15 represents the relative importance of the kinematic properties to the classification of the neural network. The 'Mother Transverse Momentum' property was the most important into separating the signal from background. On the violin plot in Fig. 13 we can observe that indeed, this property seems to distinguish the signal quite well for this particular decay ($\pi^{\pm}\mu^{\mp}$). Another important kinematic property is the Impact Parameter, also visible from the violin plot.

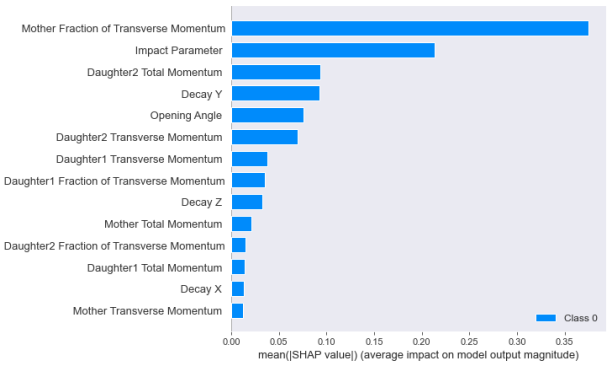

Figure 15: Relative importance of each kinematic property for the neural network, for the $\pi^{\pm}\mu^{\mp}$ decay (using SHAP)

Since we want to fully separate signal from background, we (often) must move the threshold parameter in order to achieve this. The area under the ROC curve was 1.00 for the considered neutralino decays.

Table 4: Classification metrics for all the decays of the neutralino.

Decay	Background	Threshold	Signal Efficiency
$K^{\pm}\mu^{\mp}$	92	0.5	99.92%
$K^{\pm}\mu^{\mp} \to \mu^{\pm}\mu^{\mp}$	433	0.59	99.89%
$K_S^0 \to \pi^{\pm}\pi^{\mp}$	1143	0.98	99.21%
$K_S^0 \to \pi^{\pm}\mu^{\mp}$	1380	0.99924	96.19%

With this method we were able to achieve really high signal efficiencies while eliminating all background. The classification task was made easier to the neural network by pre-processing the data distributions, resulting in feature information that facilitates the classification task.

6 Conclusions

By using the SHiP software framework we were able to simulate Hidden Sector particles (signal) as well as neutrinos (noise). By capturing kinematic properties of the signal and background we were able to develop methods to distinguish between these two: cut-based selection, Neural Network having the kinematic properties as input, and more advanced machine learning methods with preprocessing of the kinematic properties probabilities.

While the first method is the easiest to implement, the results are sub-optimal, giving signal efficiencies of 0.60-0.70; the second method is better, with a signal efficiencies up to 0.80-0.90; the performance of the last method is by far the highest, giving signal efficiencies of 0.95-1.00.

The next steps would be to simulate data with different hidden particles, in order to explore additional aspects of the hidden sector of particle physics. Also studying more realistic detector conditions, and testing the effect of varying levels of vacuum contamination. Towards achieving a more refined and extended characterization of the sensitivity of the SHiP experiment. This work demonstrates that significant improvements may be achieved through the exploration of advanced machine learning methods in the exploration of signals of New Physics.

7 Acknowledgments

We would like to thank LIP (Laboratório de Instrumentação e Física Experimental de Partículas) for providing these summer internships, an unique opportunity. This paper and the research behind it would not have been possible without the exceptional support and guidance of Nuno Leonardo and Guilherme Soares throughout this internship. Celso Franco was also involved in the earlier preparation of the internship project.

References

[1] SHiP Collaboration, *A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case*, arXiv:1504.04855v1 (2015)

[2] J. Feng et al. (2014), 1401.6085

[3] M. Cristinziani, *The SHiP experiment at CERN*, arXiv:2009.06003v1 (2020)

[4] A. Blanco, F. Clemencio, P. Fonte, C. Franco, N. Leonardo, L. Lopes, C. Loureiro, J. Saraiva, G. Soares, Journal of Instrumentation **15**, C10017 (2020)

[5] M. Al-Turany, *The fairroot framework*, M Al-Turany et al 2012 J. Phys.: Conf. Ser. 396 022001 (2012)

[6] SHiP Collaboration, *Sensitivity of the SHiP experiment to Dark Photons* (2020)

[7] SHiP Collaboration, *A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case* (2015), chap. 6.2

[8] C. Bini, *Data analysis in particle physics*, StatEPP (2014)

[9] P.C. Bhat, *Advanced analysis methods in particle physics*, FERMILAB-PUB-10-054-E (2010)

[10] F. Safara, R. Santos, G. Soares et al. (2020), SHiP-Distribution-Analysis