Hit Source	Class	MC Fraction (%)	Data Fraction (%)	Data / MC
All	Track	77.1	76.2	0.99
Pion	Track	93.8	95.4	1.02
Proton	Track	97.1	96.6	0.99
Electron	Shower	96.9	95.8	0.99
Cosmic–ray	Track	92.9	90.3	0.97

Table 6.4: Fraction of hits in correct class for different samples in ProtoDUNE–SP data and simulation.

the numbers in each case is an estimate of the percentage uncertainty associated with this simple selection algorithm. For the sample of all hits, the track category was used, and a fractional difference of 1% was seen. This varies depending on the particle species, with the largest difference being 3% for the cosmic–ray sample.

Overall, the CNN results have been shown to agree well with data across a range of particle types. The discrepancies seen are also influenced by the difference between data and simulation for the Pandora reconstruction framework, which makes the cause of the discrepancies difficult to disentangle. However, the general agreement between data and simulation is a sign that the results from the CNN are sensible, and the additional classification strength of the CNN over Pandora makes it a useful tool in analyses of the ProtoDUNE–SP data. The discrepancies will impact each analysis differently, therefore, the uncertainties involved with using the CNN classifier should be evaluated on a case–by–case basis; for hit selection with a simple hit–by–hit algorithm the uncertainties are on the order of 1-3% depending on the particle species.

6.4 Application in ProtoDUNE–SP Analyses

The output of the CNN classifier has been applied in a number of ProtoDUNE–SP analyses since it was developed. Chapter 7 will detail one use of the network in a Michel electron analysis. In addition, the scores from the network are being incorporated into analyses by other members of the ProtoDUNE–SP experiment, including:

- Selecting shower candidates for neutral–pion event selection[115].
- Identifying charge exchange candidates in charged–pion cross section analyses[116].
- Identifying Michel electron contaminated tracks for stopping muon calibration[117].

In the neutral pion analysis, the CNN score is being used as part of the sample selection algorithm, which is used to select a sample of events with pairs of secondary photons. An example of the CNN shower score distribution for secondary particles in this sample is shown in Figure 6.15, demonstrating the breakdown of the CNN score by particle species for candidates in this sample. An example of a neutral pion candidate selected with this algorithm is shown in Figure 4.5d in Chapter 4. These plots are courtesy of Milo Vermeulen, who developed the neutral pion event selection algorithm in ProtoDUNE–SP[115].

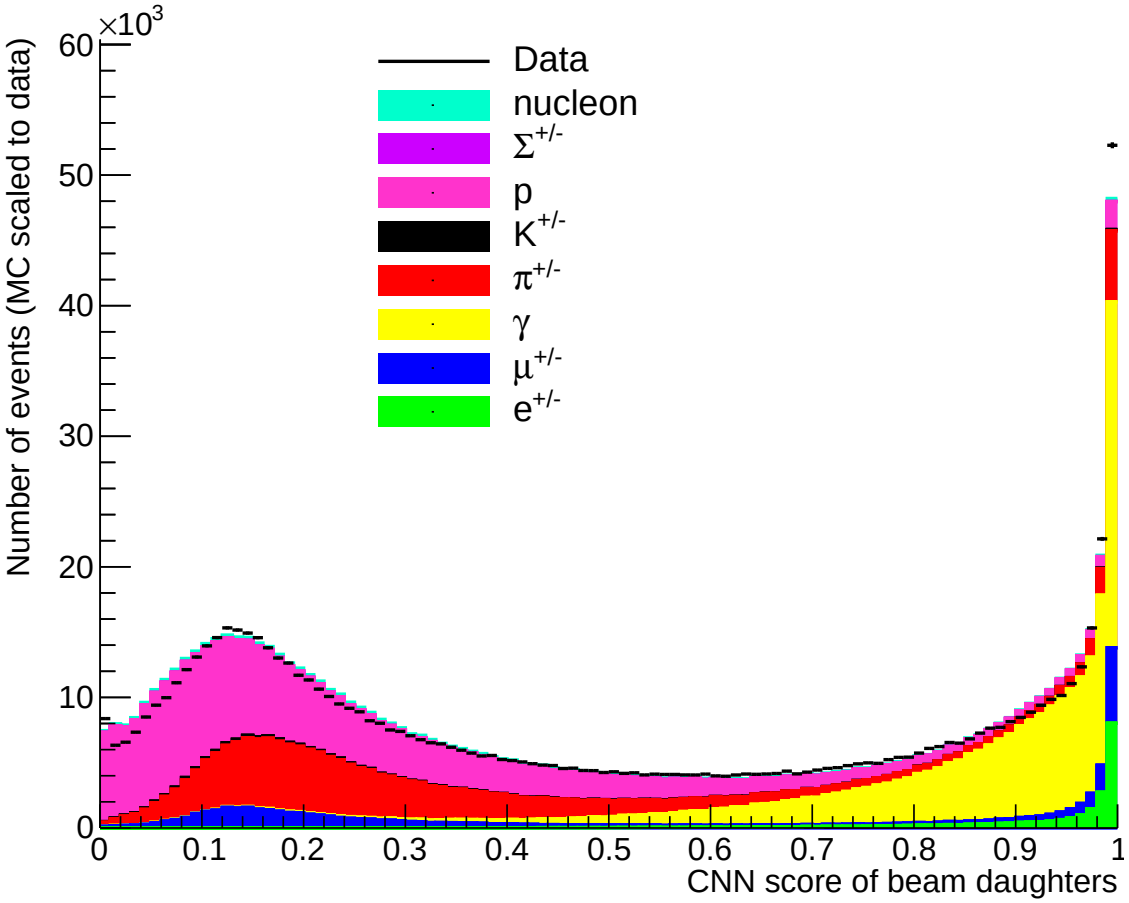

Figure 6.15: CNN shower score distribution for secondary particles in neutral pion event selection sample. The coloured distributions represent different true particle species in ProtoDUNE–SP simulation, and the black points represent the distribution in ProtoDUNE–SP data. Figure courtesy of Milo Vermeulen[115].