
Machine Learning nella Fisica delle Alte Energie

Sono stato colpito dall'urgenza di fare. Conoscere non è abbastanza; dobbiamo applicare. Essere disposti non è abbastanza; dobbiamo fare.

Leonardo da Vinci

Konstantinos Bachas

Aristotle University of Thessaloniki, Thessaloniki, Greece

Stefania Spagnolo

*Dipartimento di Matematica e Fisica "E. de Giorgi", Università del Salento
Istituto Nazionale di Fisica Nucleare, sez. di Lecce*

Intelligenza Artificiale e *Machine Learning* (ML) hanno una relazione di vecchia data con la Fisica delle Alte Energie. Le tecniche di *Machine Learning* hanno giocato un ruolo importante nell'analisi dei dati della Fisica delle Alte Energie nelle due ultime decadi. Oggi, questa relazione è anche più stretta nel contesto della prossima era del Large Hadron Collider (LHC) del CERN e i suoi esperimenti. LHC rappresenta lo strumento principe della ricerca nella fisica delle alte energie e, nello stesso tempo, una sfida che motiva lo sviluppo di metodologie potenti per sfruttare al massimo l'enorme ammontare di dati prodotti nella loro complessità, alta dimensionalità e numerosità. In aggiunta, LHC è l'esempio che illustra al meglio perché la fisica delle alte energie è sempre stata un terreno fertile per innovazioni di riferimento per la gestione dati e *Machine Learning*. In questo articolo presenterò qualche esempio di applicazio-

Artificial Intelligence and Machine Learning (ML) have a long-standing relationship with High Energy Physics. Machine Learning techniques have played an important role in the analysis of high-energy physics data over the last couple of decades. Today, this relationship is even tighter in the context of the ongoing era of the Large Hadron Collider at CERN and its experiments. The latter constitute the state-of-the-art tools to perform research in high-energy physics and at the same time a challenge to find powerful ways to exploit as much as possible the huge amounts of produced data due to their complexity, high-dimensionality and size. In addition, the Large Hadron Collider is also a very illustrative example why high-energy physics has always been an excellent ground for bench-marking innovations in data science and Machine Learning. In this article, examples of application of Artificial Intelligence and

ne dell'Intelligenza Artificiale e di tecniche di *Machine Learning* nell'analisi dei dati della Fisica delle Alte Energie al Large Hadron Collider, che dimostra la potenzialità del *Machine Learning* nell'affrontare difficili problemi di analisi dati e nel fornire soluzioni promettenti alla sfida proposta dai moderni esperimenti di fisica delle particelle elementari.

Introduzione

La Fisica delle Alte Energie (HEP) è la branca della scienza che studia le leggi fondamentali che governano la natura e le interazioni delle particelle elementari.

La nostra comprensione attuale della fisica delle alte energie è condensata nel quadro teorico denominato Modello Standard (SM). Lo SM è una teoria quantistica di campo che è stata sviluppata più di 50 anni fa e incorpora sia la meccanica quantistica che la relatività speciale.

Nonostante il suo enorme successo, lo SM non è la teoria definitiva. Alcune questioni fondamentali non trovano spiegazione nello SM e, allo stesso tempo, gli esperimenti condotti finora non hanno fornito nessuna indicazione utile alla loro soluzione. Ciò suggerisce che ci sia una teoria più fondamentale. Ci sono diversi esempi impressionanti di queste notevoli questioni. La differenza nell'intensità tra le forze elettrodeboli e la gravità, l'origine e la composizione della materia oscura, e l'apparente asimmetria tra materia e anti-materia nell'universo. Rispondere a queste domande ed esplorare in dettaglio l'origine delle masse delle particelle elementari sono le ragioni principali che hanno motivato la costruzione di LHC al CERN vicino a Ginevra, Svizzera.

Il Large Hadron Collider

LHC è il più grande e più potente acceleratore di particelle del mondo. Esso fa collidere, nel suo tunnel circolare di lunghezza pari a 27 km, posto 100 m sottoterra, due fasci di protoni che si muovono su traiettorie opposte a quasi la velocità della luce e all'energia più alta mai raggiunta dall'essere umano. Fino ad ora, LHC ha funzionato eccezionalmente bene e ha prodotto una ricca messe di risultati di fisica tra i quali il più salien-

Machine Learning techniques in High Energy Physics data analysis at the Large Hadron Collider are presented, demonstrating, in yet another scientific area, the potential of Machine Learning in addressing difficult problems in data science and providing promising solutions to the challenges of modern particle physics experiments.

Introduction

High Energy Physics (HEP) is the scientific domain which studies the fundamental laws governing nature and the interactions of elementary particles.

Our current understanding of elementary particle physics is represented by the so-called Standard Model theoretical framework. The Standard Model (SM) is a quantum field theory which is being developed for more than 50 years and incorporates consistently both quantum mechanics and special relativity.

Despite its huge success, the SM is not the ultimate theory and there are still open questions that the SM cannot accommodate and have not been addressed experimentally, suggesting that there is a more fundamental theory. There are several striking examples of such outstanding questions. The difference between the strength of electro-weak and gravity forces, the origin and composition of dark matter, and the apparent matter-antimatter asymmetry in the universe. Answering these questions and fully exploring the origin of mass of the elementary particles are the main reasons for building the Large Hadron Collider (LHC) at CERN near Geneva, Switzerland.

The Large Hadron Collider

The LHC is the biggest and most powerful particle accelerator on earth, colliding, in its 100 m underground tunnel of 27 km circumference, two counter-rotating beams of protons almost at the speed of light and at the highest energy ever reached. The LHC has so far performed exceptionally well and delivered a wealth of physics results with the highlight of the Higgs boson discovery in 2012 [1, 2, 3].

te è la scoperta, nel 2012, del bosone di Higgs [1, 2, 3].

D'altra parte, LHC è anche la sorgente più abbondante al mondo di dati scientifici in termini di quantità, ma anche di complessità. I protoni dei fasci che circolano nel tunnel di LHC raggruppati in pacchetti molto densi collidono nei quattro punti di intersezione delle traiettorie dei fasci (zone di interazione) ogni 25 nano secondi. Le collisioni generano il tasso di eventi, mai ottenuto prima, di 40 milioni di eventi per secondo. Ogni collisione produce un gran numero di particelle registrate dai rivelatori degli esperimenti installati attorno alle zone di interazione che legono $O(10^8)$ sensori di diversi tipi in ogni evento. Questi grandi tassi di collisioni sono necessari a causa della natura probabilistica delle interazioni e del gran numero di diversi stati finali che possono essere generati in una singola collisione. Eventi che generano prodotti interessanti sono estremamente rari.

Di conseguenza, la velocità con cui LHC produce dati supera di ordini di grandezza le moderne capacità di processare ed immagazzinare dati. Per questo motivo soltanto una piccola frazione di questi eventi è conservata per un'ulteriore analisi. Algoritmi elaborati, ma veloci, selezionano solo alcuni eventi ritenuti interessanti sulla base di criteri fisici riducendone il tasso a circa 1 kHz. Tuttavia, anche con questo fattore di riduzione, durante il periodo di acquisizione dati detto Run 1 di LHC (2009 - 2013), il tasso di scrittura dei dati su disco era di 1 gigabyte per secondo (Gb/s), con un picco occasionale di 6 Gb/s [4]. Nel run2 di LHC (2015-2018) la velocità con cui i dati sono stati prodotti e scritti su disco è diventata di 8 Gb/s; questo valore raddoppierà nel prossimo run ed infine quintuplicherà nel run di alta luminosità di LHC (a partire dal 2027) quando si arriverà ad immagazzinare 500 petabyte di dati ogni anno.

L'analisi di questo enorme volume di dati rappresenta una sfida alla gestione dei dati e quindi per fisica delle particelle. Essa impone di considerare il (Machine Learning) (ML) una metodologia decisiva per il raggiungimento degli obiettivi scientifici di LHC in quanto essa aiuta a ridurre la dimensione dei dati concentrando l'informazione rilevante contenuta in dati di basso livello e alta dimensionalità in una rappresentazione di

Seen from another point of view, the LHC is also the world's most abundant source of scientific data in terms of size but also in terms of complexity. The colliding beams at the LHC are grouped into bunches of protons which cross every 25 nsec and the proton-on-proton collisions yield an unprecedented event rate of about 40 million events per second. Each collision produces a large number of particles which are being recorded by the LHC's detectors installed at specific locations around the accelerator ring and read $O(10^8)$ sensors of different types. These high event rates are necessary because of the probabilistic nature of the collisions and the physics channels produced in each event. Events that result in interesting products are very rare.

This event rate is orders of magnitude beyond nowadays data processing and storage capabilities. Because of this, only a small fraction of the data is saved for further analysis. For each of the produced events from a proton-on-proton head-on collision, elaborate algorithms select the interesting events based on physics criteria which reduces the rate down to roughly 100 kHz. However, even with this reduction factor, during the so-called Run 1 data-taking period of the LHC (2009 – 2013), data storing rate was 1 gigabyte-per-second (Gb/s), with the occasional peak of 6 Gb/s [4].

During Run 2 (2015-2018) the data storage speed raised to 8 Gb/s. This value will double in the next run and, finally, it will increase by a factor of five for the High Luminosity run of LHC (starting from 2027). At that point, the data volume written on tape each year will be 500 petabyte.

The analysis of this huge volume of data and the fact that the data size is expected to double at the next LHC Run scheduled for 2021-2023, pose a challenging task for data science and particle physics and indicate that Machine Learning (ML) can have a decisive effect on the LHC scientific goals as it can improve data reduction, reducing the relevant information contained in the low-level, high-dimensional data into a higher-level and lower-dimensional space.

alto livello e bassa dimensionalità.

I rivelatori degli esperimenti a LHC

L'esperimento ATLAS [5] è uno dei due grandi rivelatori *general-purpose* progettati per sfruttare l'opportunità di indagine sulle particelle elementari e le loro interazioni ad energie della decina di TeV offerta da LHC. L'apparato è installato in una caverna sotterranea a 100 m dal suolo, ha una forma cilindrica, una copertura a 4π e le sue dimensioni sono di 25 m di altezza e 46 m di lunghezza; il peso non inferiore a $7 \cdot 10^3$ tonnellate è simile a quello della torre Eiffel. ATLAS consiste di sei diversi sottosistemi di rivelazione avvolti come strati di forma cilindrica attorno al punto di collisione con la funzione di registrare il passaggio di ogni particella prodotta nelle collisioni misurandone il momento e l'energia e, in alcuni casi, identificandone la natura. Probabilmente è uno degli apparati più complessi mai costruiti.

L'esperimento CMS [6] è un rivelatore *general-purpose* a LHC con gli stessi obiettivi scientifici di ATLAS. Il rivelatore CMS è costruito attorno ad un enorme magnete solenoidale. Il rivelatore completo è lungo 21 m, largo 15 m, alto 15 m e ha un peso di circa $14 \cdot 10^3$ tonnellate.

I rivelatori ATLAS e CMS, in ogni secondo, sono spettatori di oltre un miliardo di interazioni tra particelle di altissima energia.

La necessità di *Machine Learning* nella fisica delle alte energie

Nonostante, comunemente, si abbia l'impressione che ML sia una tecnica utilizzata da pochi anni nella fisica delle particelle elementari, in realtà, essa ha giocato un ruolo importante nell'analisi dei dati per decenni. Tuttavia, solo con il programma di LHC la fisica delle particelle ha prodotto un così vasto e complesso volume di dati da rendere il ML una risorsa di utilizzo comune e diversificata. I dati registrati dagli esperimenti ATLAS e CMS sono infatti molto numerosi e ad alta dimensionalità.

Diversi approcci di ML sono stati applicati per affrontare un'ampia varietà di problemi in HEP, dalla classificazione degli eventi, alla ricostruzione delle tracce nei rivelatori, all'identificazione di alcune particelle come elettroni, fotoni o lep-

The detector experiments at the LHC

The ATLAS experiment [5] is one of the two large general-purpose detectors designed to exploit the physics potential of the LHC. It is installed in a cavern 100 m below ground, it has cylindrical shape and 4π coverage and dimensions of 25 m height and 46 m length while it weights not less than $7 \cdot 10^3$ tonnes, similar to the weight of the Eiffel Tower. It consists of six different detecting subsystems wrapped concentrically in layers around the collision point to record the trajectory, momentum, and energy of particles, allowing them to be individually identified and measured and is arguably one of the most complex devices ever built.

The CMS experiment [6] is a general-purpose detector at the LHC with the same scientific goals as the ATLAS experiment. The CMS detector is built around a huge solenoid magnet. The complete detector is 21 metres long, 15 metres wide and 15 metres high and has a weight of $14 \cdot 10^3$ tonnes

Over a billion particle interactions take place in each of the ATLAS and CMS detectors every second.

The need for Machine Learning in HEP

Although the impression is that applications of ML in HEP is something which is going on for just a few years, in fact, ML has played an important role in the physics data analysis for decades. However, not before the LHC program commenced had a particle physics experiment provided with such complex and big volumes of data. The data recorded by the ATLAS and CMS experiments are innumerable and high-dimensional.

Several ML approaches have been applied to tackle a wide variety of problems in HEP, from event classification, detector hit reconstruction, to object identification and reconstruction by using information from various detector systems, such as electron, photon, or τ lepton identification or even to identify interesting collision

toni τ mediante l'uso combinato di informazioni provenienti da vari sistemi di rivelazione. Infine il ML è utilizzato anche in algoritmi di selezione in tempo reale per l'identificazione di processi interessanti. L'uso di gran lunga più frequente del ML è avvenuto nell'ambito del problema di classificare eventi come di rumore o di segnale.

Tradizionalmente, i fisici delle alte energie usano tecniche di analisi e di riduzione dei dati che consistono in sequenze di selezioni binarie su varie proprietà dell'evento e, successivamente, costruiscono per gli eventi selezionati la distribuzione di una singola quantità che è poi processata con una procedura statistica. Questo approccio all'analisi è spesso chiamato *cut-based analysis*. Ad esempio, per identificare l'esistenza o meno di una nuova particella prevista da un modello fisico teorico (o anche per una particella esistente in un processo noto nello SM) si cerca un sottoinsieme dei dati (ad alta dimensionalità) da collisione protone-protone per una loro analisi successiva basata sull'informazione associata a ciascuno degli eventi selezionati. L'obiettivo è ottenere sotto-campioni selezionati in modo che gli eventi dovuti a processi che producono la nuova particella - il segnale - compaiano in numero statisticamente significativo rispetto ad altri eventi - il fondo - che hanno caratteristiche simili al segnale stesso. All'ipotesi di presenza del segnale in queste regioni dei dati dovrebbe quindi corrispondere una previsione significativamente differente rispetto a quella relativa all'ipotesi nulla (assenza di segnale), permettendo, quindi, una solida verifica statistica [8].

I criteri di selezione dei dati sono guidati principalmente da considerazioni ed intuizioni fisiche che non sono facili da estendere ad alte dimensionalità e non c'è alcuna garanzia che queste selezioni siano le più efficienti. ML viene utilizzato per affrontare questi problemi fornendo una riduzione della dimensionalità e una migliore prestazione rispetto agli approcci tradizionali.

Fino ad ora, tra i numerosi algoritmi di Intelligenza artificiale, quelli che sono stati utilizzati più frequentemente per l'analisi *off-line* di dati di fisica di particelle elementari, sono le architetture semplici come *boosted decision trees* (BDT) o *shallow Neural Networks* (NN). Oggi, invece, le applicazioni di *deep neural networks* (DNN) che hanno la potenzialità di gestire problemi di alta

events and provide triggering for the detectors to record them. Most often though ML tools have been used to classify entire events as background-like or signal-like.

Traditionally, high energy physicists use data analysis and data reduction techniques utilising sequences of binary selections on the distribution of a single observed quantity, followed by a statistical treatment of the selected data. This is often called with the term **cut-based analysis**. For example, to search for the existence or not of a new particle predicted by a theoretical physics model (or even for an existing particle process known in the SM), one tries to extract subsets of the high-dimensional collision data for further analysis based of the information associated with individual events. The ultimate goal is to select these subsets in such a way that events due to processes producing the new particle -the signal- have statistical significance over other events-the background- from processes that mimic the signal ones. The signal hypothesis in these data regions would then exhibit significantly different predictions than the null hypothesis, thus allowing for an effective statistical test [8].

The decisions on the selection of the data are driven mainly by physics considerations and insights, which are not easily extended to higher dimensions nor there is any guarantee that the selections are the most efficient ones. ML comes to address these problems, by providing dimensionality reduction and improved performance with respect to the traditional approaches.

Out of the plethora of the Artificial Intelligence algorithms, simple architectures such as boosted decision trees (BDT) or shallow Neural Networks (NN), have been most frequently used in the off-line HEP analysis, while applications of deep neural networks (DNN) which could adeptly handle higher-dimensional and more complex problems than previously feasible, are now more

dimensionalità più complessi rispetto a quelli affrontabili in precedenza, riscuotono sempre più interesse, anche se il loro uso non è ancora così ampiamente diffuso. Tecniche di ML sono già state applicate a varie analisi condotte sui dati degli esperimenti ATLAS e CMS qui in seguito verranno presentate due applicazioni interessanti: una riguardante l'uso di una NN ricorsiva e l'altra basata su una DNN parametrica.

Applicazione di *recurrent Neural Networks* nella ricerca di risonanze pesanti con il rivelatore ATLAS a LHC

Molte estensioni dello SM prevedono l'esistenza di nuove particelle pesanti (risonanze) che decadono in coppie di bosoni (dibosoni) chiamati W e Z che, a loro volta, decadono in quark o leptoni (elettroni, muoni e tau, oltre agli sfuggenti neutrini). La produzione di queste coppie avviene dopo la collisione di due protoni e, più precisamente, per il tramite dell'interazione di due dei costituenti del protone, due gluoni o due quark. A seconda del modello teorico di riferimento, il dibosone previsto può essere prodotto da tre meccanismi distinti: la fusione gluone-gluone (ggF), il processo di Drell-Yan (DY) oppure la fusione vettore-bosone (VBF).

and more of gaining interest, even if they are not yet widely spread in this domain. ML techniques have been applied already to several physics analyses by both the ATLAS and CMS experiments and two interesting applications of recurrent NN and parameterized DNN are presented in the following.

Applications of recurrent Neural Networks in searches for heavy resonances with the ATLAS detector at the LHC

Many extensions to the SM predict the existence of heavy new particles (resonances) that decay into pairs of bosons (dibosons) named W and Z which in turn decay to quarks or leptons (electrons, muons, taus, besides the elusive neutrinos). The production of such pairs results from the collision of two protons and more specifically proceeds through the interaction of two of the constituents of the proton, either two gluons or two quarks. Depending on the underlying theoretical model, the predicted diboson can be produced through three distinct mechanisms namely gluon-gluon fusion (ggF), Drell-Yan (DY), or vector-boson fusion (VBF).

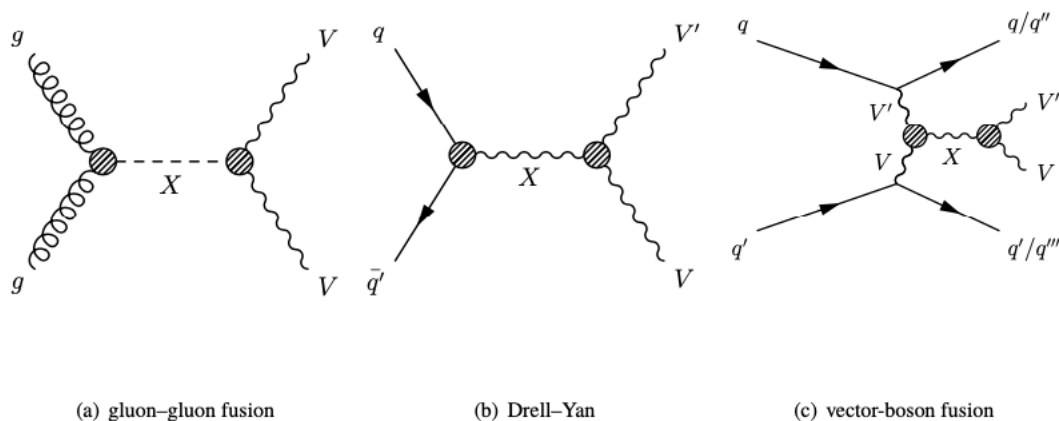


Figura 1: Diagrammi di Feynman che rappresentano la produzione di nuove particelle pesanti indicate con X con i loro decadimenti in una coppia di bosoni, indicati collettivamente come V : (a) fusione gluone-gluone, (b) Drell-Yan, (c) fusione vettore-bosone. I cerchi tratteggiati rappresentano gli accoppiamenti di X con le altre particelle, sia diretti che effettivi.

Representative Feynman diagrams for the production of new heavy particles denoted as X with their decays into a pair of bosons, denoted collectively as V : (a) gluon - gluon fusion, (b) Drell - Yan, (c) vector-boson fusion. The hashed circles represent direct or effective couplings of X to other particles.

Una rappresentazione grafica della fisica sottostante questi processi è data dai diagrammi di Feynman¹ che sono mostrati in Figura 1 per i tre meccanismi di produzione.

Come si può vedere nei diagrammi della Figura 1, dei tre processi di produzione, il ggF e il DY hanno gli stessi stati finali, mentre il VBF mostra la presenza di due quark addizionali nello stato finale. Bisogna notare che per la caratteristica della Cromo - Dinamica - Quantistica (QCD) i quark non sono mai osservati come particelle libere, ma producono un getto di particelle energetiche (adroni), che sono rivelati individualmente dall'apparato sperimentale e successivamente sono aggregate per ricostruire un'unica entità, un getto (*jet*). Nel caso di modalità di produzione VBF i due quark addizionali nello stato finale sono indicati come VBF *jet*. Quest'ultima caratteristica rappresenta una differenza chiave con le altre due modalità di produzione, ggF e DY, poiché la cinematica dei *jet* VBF differisce da quelle dei *jet* generati dal decadimento dei bosoni Z e W in quark. Tipicamente i VBF *jet* sono ben separati in **pseudorapidity** e la loro massa invariante è grande.²

Queste proprietà cinematiche sono state usate in ricerche precedenti a LHC per separare processi di produzione VBF da processi di produzione in ggF/DY. In una pubblicazione recente [9] dell'esperimento ATLAS è stata, invece, usata una *recurrent neural network* (RNN) [10] per discriminare tra i meccanismi di produzione VBF e ggF/Dy. Questa classificazione del meccanismo di produzione basata su un RNN avviene all'inizio del flusso di analisi che prosegue con una successione di tagli secondo la tradizione di una *cut-based analysis* conduce alla decisione finale riguardo all'ipotesi che l'evento sia prodotto in un processo di segnale o di fondo. La referenza [9] presenta la ricerca di risonanze pesanti che decadono in WW , ZZ or WZ nella collisione protone - protone all'energia del centro di massa di 13 TeV ed è effettuata per stati finali in cui un bosone W o Z decade in leptoni mentre l'altro bosone W o Z decade in due *jet*.

¹I diagrammi di Feynman sono un elemento essenziale del linguaggio della fisica delle particelle essendo un mezzo potente per illustrare e calcolare le transizioni tra stati nella teoria quantistica dei campi.

A pictorial representation of the fundamental physics underlying these processes is provided by the Feynman diagrams¹ which are shown in Figure 1 for the three production mechanisms.

As it can be seen in the diagrams in Figure 1, from the three production processes, the ggF and DY processes have the same final states while the VBF process exhibits two additional quarks in the final state. It should be noted that owing to the nature of Quantum Chromo-Dynamics (QCD) quarks are never observed as free particles, but they always give rise to a jet of energetic particles (hadrons), which are detected by the experimental apparatus and reconstructed as an individual object, a jet. In the case of the VBF production mode the two additional quarks in the final state are called VBF jets. The latter is a key difference with respect to the other two production mechanisms, ggF and DY, since the kinematics of the VBF-jets differ from those jets emerging from the Z or W boson decays to quarks. They are typically well separated in pseudorapidity² and usually has large dijet invariant mass.

These characteristics were used in previous searches at the LHC to separate VBF production from ggF/DY production. In a recent publication from the ATLAS experiment [9], a recurrent neural network (RNN) [10] is used to discriminate between the VBF and ggF/DY production mechanisms. This RNN based classification of the production mechanism is performed early in the analysis flow and once made, a "cut-based" analysis flow follows which provides the final decision of whether the event is a signal event or a background event. Reference [9] reports on a search for heavy resonances decaying into WW , ZZ or WZ using proton-proton collision data at a centre-of-mass energy of 13 TeV and is performed for final states in which one W or Z boson decays to leptons while the other W boson or Z boson decays to two jets.

¹Feynman diagrams are an essential part of the language of particle physics as they are a powerful way to illustrate and calculate the transitions between states in quantum field theory.

²The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$.

La motivazione per l'uso di questa architettura ML risiede nel fatto che le RNN permettono di processare sequenze di dati di lunghezza variabile. Esse possono quindi essere utilizzate per descrivere un evento in termini delle proprietà dei suoi *jet*, il cui numero varia da un evento all'altro. Alla descrizione dell'evento potrebbero inoltre contribuire utilizzando le proprietà delle altre particelle dello stato finale, elettroni, muoni o qualsiasi altra particella, che sia stata rivelata. Dato che il numero di particelle di ciascun tipo varia in ogni evento, la flessibilità dei RNN risulta necessaria per processare questo tipo di informazione.

In questo articolo della collaborazione ATLAS, il classificatore RNN usa l'informazione dei quadri-impulsi dei *jet* come variabili di input. La RNN è costruita con la libreria Keras [11] utilizzando la libreria python Theano [12] come *back-end* per i calcoli matematici. Lo studio considera tre diversi modelli teorici che descrivono la produzione di nuove particelle pesanti che decadono in due bosoni. La sottostante distribuzione delle variabili utilizzate come *input* dal RNN e, quindi, la distribuzione di probabilità del risultato del classificatore dipendono dal modello utilizzato per descrivere la risonanza pesante, la sua massa e il modo di decadimento.

²La *pseudorapidity* è definita in termini dell'angolo polare θ come $\eta = -\ln \tan(\theta/2)$.

L'addestramento della RNN viene attuato utilizzando dati di fisica simulati con metodi Monte Carlo in cui una particella pesante di massa 1 TeV decade in due bosoni Z che poi decadono in due elettroni o muoni di carica opposta e due *jet* attraverso il processo che può essere riassunto in $X \rightarrow ZZ \rightarrow \ell\ell qq$, dove ℓ denota gli elettroni o muoni e q denota i quark e, in definitiva, i *jet* ricostruiti.

In Figura 2 si mostra la distribuzione di probabilità della risposta del classificatore ad eventi simulati VBF e ggF/DY corrispondenti alla produzione di una risonanza di 1 TeV nei tre diversi modelli teorici di segnale considerato in questa ricerca. Valori della risposta prossimi a 1 sono molto probabili per processi VBF, mentre quelli vicini a 0 sono tipici di eventi ggF/DY.

La decisione sulla classificazione in base al meccanismo di produzione di ogni evento è quindi fissata da una soglia a 0.8 sulla risposta della

The motivation for the use of such a ML architecture lies in the fact that RNNs allow to process variable length sequences of data. It can therefore be used to describe an event in terms of the properties of its jets, whose number varies event by event. The event could as well be described using the properties of other final state particles which are reconstructed in the detector such as muons, electrons, or any other particle that appears in it. Because the number of particles of each type changes in each event, the flexibility of RNNs is needed to process such type of data.

In this paper by the ATLAS collaboration, the RNN uses information of the four-momenta of jets as input features. The RNN is built with the Keras [11] library using the Theano python library [12] as a back-end for the mathematical computations. The study considers three theoretical physics models that provide the production of a heavy new particle decaying to two bosons. The underlying distributions of the input features to the RNN and therefore the probability output of the classifier depends on the assumed model of a heavy resonance, its mass and on the decay mode.

The training of the RNN is preformed using Monte Carlo simulated physics data in which a heavy new particle of mass 1 TeV decays to two Z bosons and in turn the latter decay to two electrons or muons and two jets through the process which can be summarized as $X \rightarrow ZZ \rightarrow \ell\ell qq$, where ℓ denotes the electrons or muons and q denotes the quarks and therefore the reconstructed jets.

The output probability of the RNN classifier is shown in Figure 2 for simulated events from VBF and ggF/DY production of a 1 TeV resonance in the signal models considered in this search. The entries that are close to 1 in the score distribution are characterized as VBF-like events while the ones close to 0 are classified as ggF/DY-like events.

The actual decision on the type of production mechanism of each event is then imposed by a threshold at 0.8 on the resulting probability of

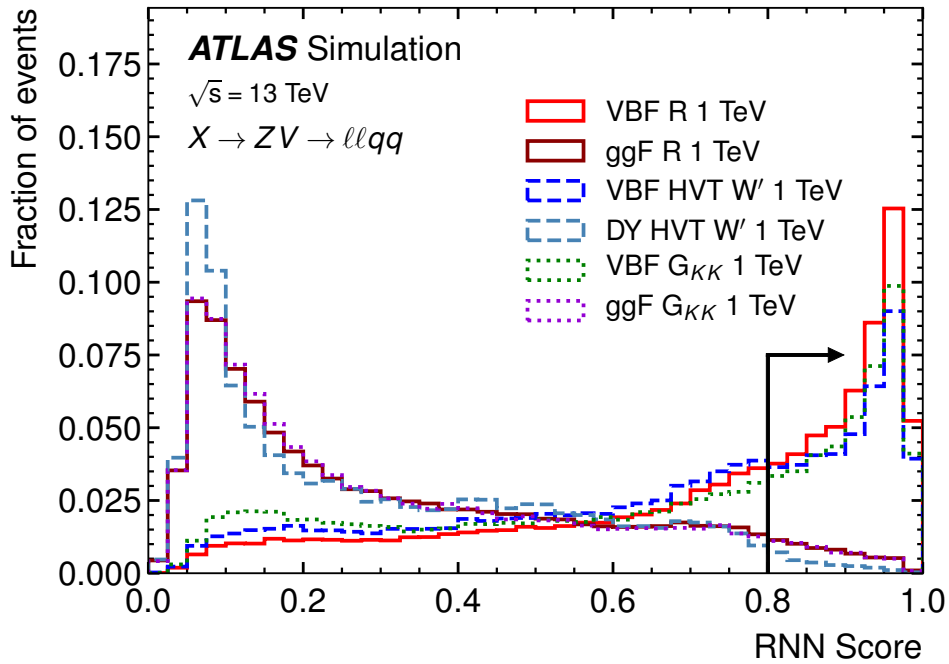


Figura 2: Distribuzione dei risultati del classificatore RNN per la produzione di una risonanza di 1 TeV nel segnale per i modelli teorici considerati nello studio pubblicato da ATLAS [9].
 The RNN classifier score distributions for the production of a 1 TeV resonance in the signal for the theoretical physics models considered in the ATLAS publication search [9].

RNN che definisce l'evento come prodotto attraverso il meccanismo VBF o come prodotto da ggF/DY. La scelta di usare questo particolare valore di soglia è conseguente ad un'analisi dedicata in cui i valori di soglia sono stati variati cercando quello che massimizza la sensibilità ad eventi di segnale prodotti attraverso VBF.

Un'idea sulla qualità della prestazione del classificatore RNN può essere estratta dalla Figura 3 che mostra la frazione di eventi di segnale simulati che superano la soglia di 0.8 sulla risposta del RNN in funzione della massa risonante per diversi modelli di segnale. Il RNN riesce a classificare correttamente gli eventi VBF più del 40% delle volte per una risonanza dibosonica di massa superiore a 1 TeV. È opportuno osservare che l'efficienza così ottenuta migliora il risultato precedente, basato su analisi *cut-based* dal 10% al 60% (a seconda del modello di risonanza e della massa della risonanza.)

the RNN, which defines the event as an event produced through the VBF or as an event which is produced through the ggF/DY mechanisms. The motivation to use this particular threshold value is driven by a dedicated analysis which scans the threshold values looking for the one that maximizes the sensitivity to VBF signals.

An idea about the performance of the RNN classifier can be obtained from Figure 3 which shows the fractions of simulated signal events passing the RNN requirement as functions of the resonance mass for different signal models. The RNN achieves to correctly classify VBF events more than 40% of the time for a diboson resonance heavier than 1 TeV. It is remarkable that this achieved efficiency is better by 10% up to 60% (depending on the resonance model and mass) than previous cut-based selection studies.

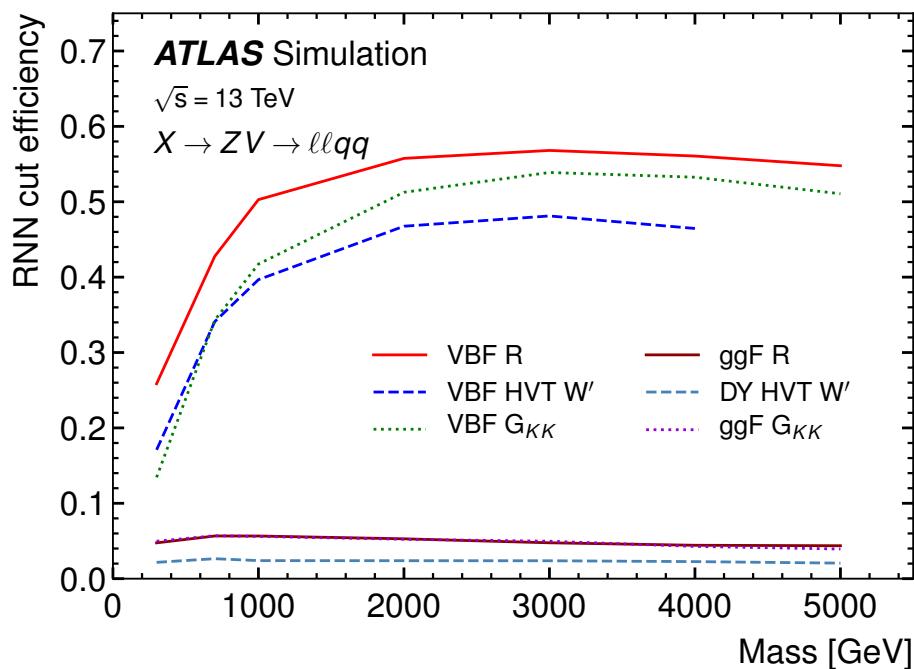


Figura 3: La frazione degli eventi di segnale che hanno superato la soglia applicata alla risposta del classificatore RNN per definire la probabilità che l'evento sia prodotto con meccanismo VBF in funzione della massa della nuova risonanza per eventi con produzione VBF e ggF. Gli istogrammi sono estratti dalla referenza [9].
The fractions of signal events passing the threshold on the RNN classifier probability for an event to be produced through the VBF mechanism as functions of the new resonance mass for both VBF and ggF production. Histogram from reference [9].

Applicazioni di DNN parametrizzate per la ricerca della produzione di coppie di bosoni di Higgs con il rivelatore CMS

Come menzionato nelle sezioni precedenti, tecniche di ML e di reti neurali sono state utilizzate nella fisica delle alte energie per affrontare, tra gli altri, problemi di classificazione. Nella sua più semplice e più frequente applicazione si tratta del problema di classificare eventi come compatibili con processi di fondo o di segnale. Questi ultimi prevedono la produzione di una particella ipotizzata teoricamente ma per la quale la teoria non riesce a prevedere il valore esatto della massa, ma può solo indicare un intervallo di valori possibili. Un esempio è stato presentato nella sezione precedente; un'estensione dello SM prevede una nuova particella X con una massa M_X . In accordo con certe ipotesi X decade in due Z che a loro volta decadono in $pp \rightarrow X \rightarrow ZZ \rightarrow \ell\ell qq$, dove p denota i protoni che collidono. All'interno

Applications of parameterized DNNs in searches for Higgs boson pair production with the CMS detector

As mentioned in the previous sections, Machine Learning techniques like Neural Networks have been used in HEP to address, among others, classification problems. In its simplest and more frequently met form is the signal-background classification problem where the signal is a theoretically predicted particle for which however the theory cannot predict the value of its mass but only provides a range of possible mass values. An example has already been given in the previous section; an extension to the SM theory predicts a new particle X with mass M_X . According to the hypothesis, X decays to two Z bosons which in turn decay to two leptons and two jets as $pp \rightarrow X \rightarrow ZZ \rightarrow \ell\ell qq$, where p denotes the colliding protons. Within the SM theoretical framework there are several other existing

del quadro teorico dello SM ci sono altri processi che generano gli stessi stati finali ad esempio la produzione attraverso interazioni nucleari forti di due *jet* e un bosone Z che decade in leptoni attraverso il processo $pp \rightarrow Z(\ell\ell) + jet$ e sono quindi considerati come fondo del processo di interesse.

In questi casi i compiti di classificazione segnale/fondo e di determinazione della massa sono distinti, ma il primo dipende dall'esito del secondo. In prima istanza per affrontare questo problema è possibile addestrare un insieme di reti neurali isolate corrispondenti ciascuna a un diverso scenario di massa. Queste reti neurali non incorporano la conoscenza di contesto generale in cui diversi valori di massa sono possibili quindi, se si cerca di usare un classificatore, che è stato addestrato a riconoscere eventi relativi ad identificare $M_X = \alpha$, per identificare la classe di eventi con $M_X = \beta$ l'accuratezza è molto bassa. Questo perché queste reti non possono interpolare bene tra diversi valori di M_X .

Per poter affrontare questo problema, nella referenza [13] è stato proposto un nuovo approccio di rete neurale parametrica in cui una singola rete neurale affronta un insieme completo di obiettivi collegati. Ciò è ottenuto estendendo la lista delle informazioni di *input* includendo in aggiunta alle quantità ricostruite quando si addestra una singola rete, anche uno o più parametri fisici che descrivono il modello nella sua generalità, come per esempio la massa della risonanza nel caso della ricerca di una particella instabile.

Una semplice rete neurale è addestrata usando delle caratteristiche di input x che sono le variabili che descrivono l'evento, ad esempio l'impulso e l'energia dello stato finale delle particelle. Al termine del processo di addestramento, il classificatore associa il valore 1 all'insieme delle variabili x in corrispondenza delle quali è stato allenato a produrre la risposta y in modo che f sia una buona previsione di y . Si mostra nella referenza [13] che se il problema in questione fa parte di un contesto più ampio descritto da uno o più parametri θ , è possibile costruire una rete neurale classificatrice funzione di x e di θ . Per un dato insieme di input x , una rete neurale tradizionale genera un numero reale $f(x_0)$. Una rete parametrica, tuttavia, fornisce una mappa che è parametrizzata rispetto a θ e produce risultati

physics processes that also lead to the same final state e.g the production via strong nuclear interaction of two jets and one Z boson which decays to leptons though the process $pp \rightarrow Z(\ell\ell) + jets$ and are therefore considered as a background to the process of interest.

In such cases, the tasks of classifying according to signal or background and of inferring the mass of the resonance are distinct, but the former depends on the actual value of the mass. To first approach this problem is tackled by training a set of isolated NNs corresponding to different mass scenarios. These networks do not incorporate knowledge of the larger context in terms of the particle's potential mass, therefore, if one tries to use a classifier, which was trained on $M_X = \alpha$, to predict the event class of an event with $M_X = \beta$ the accuracy is very poor. This is because these networks cannot interpolate smoothly between the different M_X .

To address this problem, a new NN approach has been proposed in reference [13] of a parameterized NN in which a single NN tackles the full set of related tasks. This is realized by simply extending the list of input features to include, in addition to the reconstructed quantities used when training a single network, also one or more physics parameters describing the wider scope of the problem, as for example the mass of the resonance in the resonant search case.

A plain NN is trained using as input features, x which are variables that describe the event, e.g. the momenta and energy of the final state particles. At the end of the training process, the resulting classifier provides a mapping f of x to the target value y so that f is a good predictor of the y value. If the problem in question is part of a larger context described by one or more parameters θ it has been shown in reference [13] that a NN classifier can be built that is a function of both x and θ . For a given set of inputs x , a traditional network evaluates to a real number $f(x_0)$. A parameterized network, however, provides a mapping that is parametrized with respect to θ and give different outputs values for different θ . This concept is illustrated in Figure 4.

differenti per diversi valori di θ . Questa idea è illustrata nella figura 4.

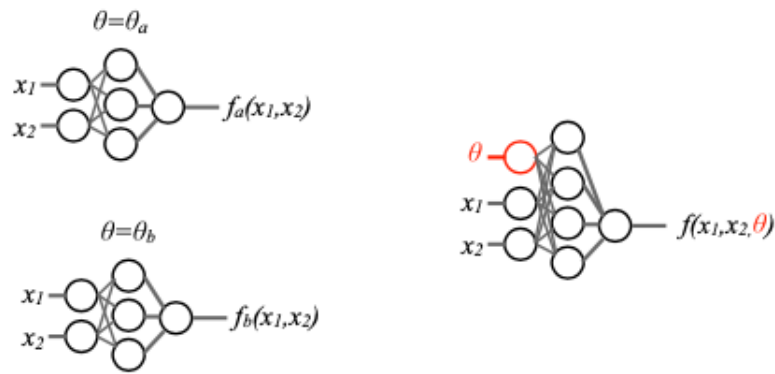


Figura 4: Illustrazione del concetto di rete parametrica, in cui invece di addestrare network individuali con caratteristiche di input (x_1, x_2, \dots, x_n) per un valore fisso del parametro di θ , un solo network è addestrato a processare le informazioni $((x_1, x_2, \dots, x_n, \theta))$. Da [13].

Illustration of the concept where instead of training individual networks with input features (x_1, x_2, \dots, x_n) for a given fixed value of parameter θ , a single network is trained with input features $((x_1, x_2, \dots, x_n, \theta))$. From [13].

Una singola rete parametrica può quindi rimpiazzare l'insieme di reti neurali individuali addestrate per specifici valori (di masse); inoltre può fornire un'interpolazione regolare della risposta in corrispondenza di valori di θ per i quali non è stata addestrata. Questa procedura semplifica anche le difficoltà tecniche relative all'applicazione delle reti neurali perché elimina la necessità di effettuare l'addestramento (che è il passaggio più dispendioso in termini di potenza di calcolo di diverse reti) ma prevede una sola rete che gestisce tutti gli scenari di masse.

L'approccio proposto in [13] ha trovato un'applicazione diretta nell'analisi fatta dalla collaborazione CMS per la ricerca della produzione risonante e non-risonante di coppie di bosoni di Higgs (HH) in scenari oltre lo SM, coppie che decadono rispettivamente in due leptoni e due neutrini ($\ell\nu\ell\nu$, dove ℓ è un elettrone, un muone o un τ), per il tramite di un bosone W o Z , e due quark di tipo b [14]. I classificatori parametrici DNN sono stati usati per migliorare la separazione del segnale dal fondo.

In particolare, l'analisi di CMS ha usato due DNN parametriche, una per la ricerca risonante, e l'altra per quella non risonante. Nel primo caso, il parametro che gioca il ruolo di θ , come descritto sopra, è la massa della risonanza ed è fornito come input al DNN, in aggiunta alle

A single parameterized network can then replace the individual NNs trained at specific (mass) points, and also smoothly interpolate to points where it has not been trained [13]. This simplifies also the technical part of the application of the NN in such physics problems because now there is no need to train several NNs (one for each mass point) but only one network for all mass scenarios is sufficient.

The approach proposed in [13] has found direct application in the analysis performed by the CMS collaboration searching for resonant and nonresonant pair-produced Higgs bosons (HH), as extensions of the SM, decaying respectively to two leptons and two neutrinos ($\ell\nu\ell\nu$, where ℓ is either an electron, a muon, or a tau lepton), through either W or Z bosons, and two b -type quarks [14]. Parameterized DNN classifiers are used to improve the signal-to-background separation.

The CMS analysis used two parameterized DNNs, one for the resonant search and the other for the nonresonant search. In the former case, the parameter that plays the role of θ , as described above, is the mass of the resonance and is provided as input to the DNNs, in addition to

caratteristiche di *input*, che sono collegate al profilo cinematico delle particelle nello stato finale. L'insieme dei parametri θ consiste di 13 possibili valori della massa compresi tra $M_X = 260$ e 900 GeV. Questi sono combinati insieme in una singola procedura di addestramento.

Nel secondo caso, i parametri θ usati sono κ_λ e κ_t : essi intervengono nella Lagrangiana che descrive l'interazione e modificano gli accoppiamenti del bosone di Higgs aumentando la produzione di coppie di bosoni di Higgs. Nel processo di addestramento del DNN sono state considerate 32 combinazioni di questi parametri con valori di κ_λ che variano tra -20 e 20 e κ_t tra 0.5 e 2.5 .

La DNN parametrica opera tanto efficacemente quanto le DNN individuali addestrate su uno specifico valore di M_X ma necessita di un singolo processo di addestramento e produce una interpolazione continua per i casi non considerati durante la fase di addestramento. Questo è illustrato nella Figura 5. Le due curve si sovrappongono e indicano che la DNN parametrica è capace di generalizzare a casi non incontrati durante la fase di addestramento interpolando il comportamento tra punti contigui di M_X .

Le distribuzioni della variabile di uscita delle DNN per produzione risonante e non risonante di una coppia di bosoni di Higgs, dopo l'applicazione di alcuni criteri di selezione preliminari, sono riportate nella figura 6. L'accordo tra dati e predizioni dello SM è evidente negli istogrammi e indica l'assenza di nuovi fenomeni.

Queste distribuzioni di risultati da DNN parametriche sono usati come discriminanti finali in [14] e possono essere usate in ogni analisi simile per ottenere limiti sulla sezione d'urto di produzione invece di altre variabili cinematiche che sono normalmente usate come le masse invarianti ricostruite del sistema di particelle. Ciò avviene grazie al loro uso come modello per un *fit* di massima verosimiglianza finalizzato all'estrazione dell'intensità del segnale di *best fit* ossia alla determinazione della sezione d'urto. Nel caso di [14], il *fit* è fatto usando modelli costruiti dalle distribuzioni dei risultati di DNN parametriche in tre regioni della distribuzione della massa invariante dei due *jet*. Questi *fit* sono quindi usati per ottenere il limite superiore al 95 % di livello di confidenza sulla sezione d'urto di produzione per X per il processo $X \rightarrow HH \rightarrow b\bar{b}VV \rightarrow b\bar{b}l\nu l\nu$,

reconstructed input features, which are related to the kinematic profile of the final state particles. The set of θ parameters are 13 possible mass values of the resonance ranging from $M_X = 260$ to 900 GeV. These were combined together in a single training.

In the second case, the θ parameters used, namely κ_λ and κ_t , are parameters that enter in the Lagrangian that describes the interaction and modify the Higgs boson couplings and enhance Higgs boson pair production. 32 combinations of these parameters are considered in the DNN training with κ_λ ranging from -20 to 20 and κ_t from 0.5 to 2.5 .

The parameterized DNN is able to perform as well as individual DNNs trained on specific M_X while requiring only a single training, and provides a smooth interpolation to cases not seen during the training phase. This is illustrated in Figure 5. Both curves overlap, indicating that the parameterised DNN is able to generalize to cases not seen during the training phase by interpolating the signal behaviour from nearby M_X points.

The distribution of the DNN output for the resonant and non-resonant production of the pair of Higgs bosons, after selection requirements, are shown in Figure 6. A nice agreement with the SM data is evident in the histograms and gives no hint of new phenomena.

Such parameterized DNN score distributions are then used as final discriminants in [14] and can be used in any similar analysis to obtain limits on the production cross section and replace other variables which are typically used like the reconstructed invariant mass of the system of particles. This is achieved by using them as templates to a binned maximum likelihood fit in order to extract the best fit signal cross sections. In the case of [14] the fit is performed using templates built from the parameterized DNN output distributions in three regions of the invariant mass distributions of the two jets. These fits are then used to obtain the 95% CL upper limits on the product of the production cross section for X and branching fraction for $X \rightarrow HH \rightarrow b\bar{b}VV \rightarrow b\bar{b}l\nu l\nu$, as a function of M_X .

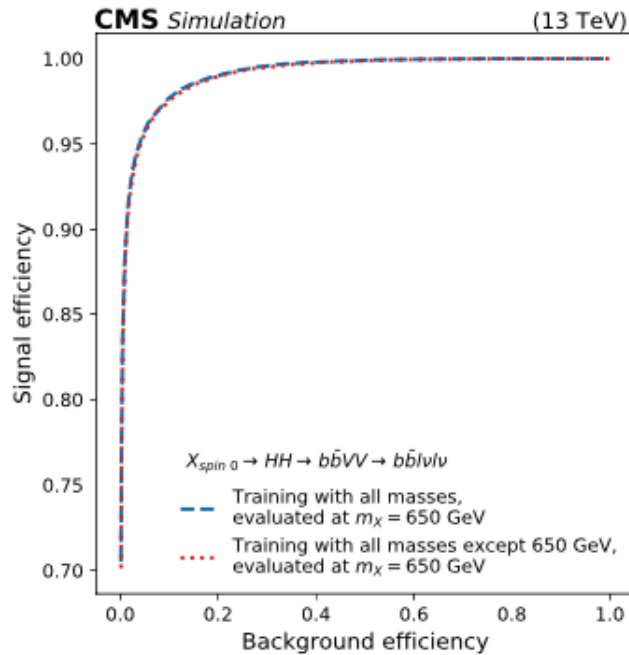


Figura 5: Curva ROC - Receiver Operating Characteristic - dell'efficienza per il segnale in funzione della percentuale di fondo accettato come segnale per una DNN parametrizzata nel caso di $M_X = 650$ GeV. La linea tratteggiata corrisponde alla DNN addestrata su tutti i possibili segnali e calcolata per $M_X = 650$ GeV. La linea a punti mostra una diversa DNN addestrata su tutti i segnali tranne quello corrispondente a $M_X = 650$ GeV, e calcolata per $M_X = 650$ GeV. Iistogrammi dalla referenza [14].

ROC curve of the signal efficiency versus background efficiency for a parametrized DNN in the case where $M_X = 650$ GeV. The dashed line corresponds to the DNN trained on all available signal samples, and evaluated at $M_X = 650$ GeV. The dotted line shows an alternative DNN trained using all signal samples except for $M_X = 650$ GeV, and evaluated at $M_X = 650$ GeV. Histogram from reference [14].

in funzione di M_X .

Conclusioni

Negli ultimi anni, i progressi nell'intelligenza artificiale e ML hanno permesso lo sviluppo di strumenti che hanno il potere di imprimere una nuova impronta alla strategia di analisi dati nella fisica delle alte energie.

Gli esempi di applicazione di reti neurali per affrontare ordinari problemi di classificazione a LHC presentati in questo articolo sono solo una minuscola frazione delle numerose aree dove intelligenza artificiale e ML sono applicate nella fisica delle alte energie. Nella grande maggioranza dei casi, questi approcci superano nelle prestazioni quelli tradizionali e hanno già un impatto fondamentale sulla metodologia dell'analisi dei dati che è, e sarà, effettuata nel futuro a LHC.

Conclusions

In the past few years, advances in Artificial Intelligence and Machine Learning have enabled the development of tools that have the power to shape the nature of data analysis in HEP.

The examples of the application of deep and recurrent neural networks to address common classification problems at the LHC described in this review, is only a tiny fraction of numerous areas where AI and ML are being applied in HEP. In the vast majority of the cases, these approaches outperform the traditional ones and have already a major impact on the way data analysis is being and will be conducted in the future at the LHC.

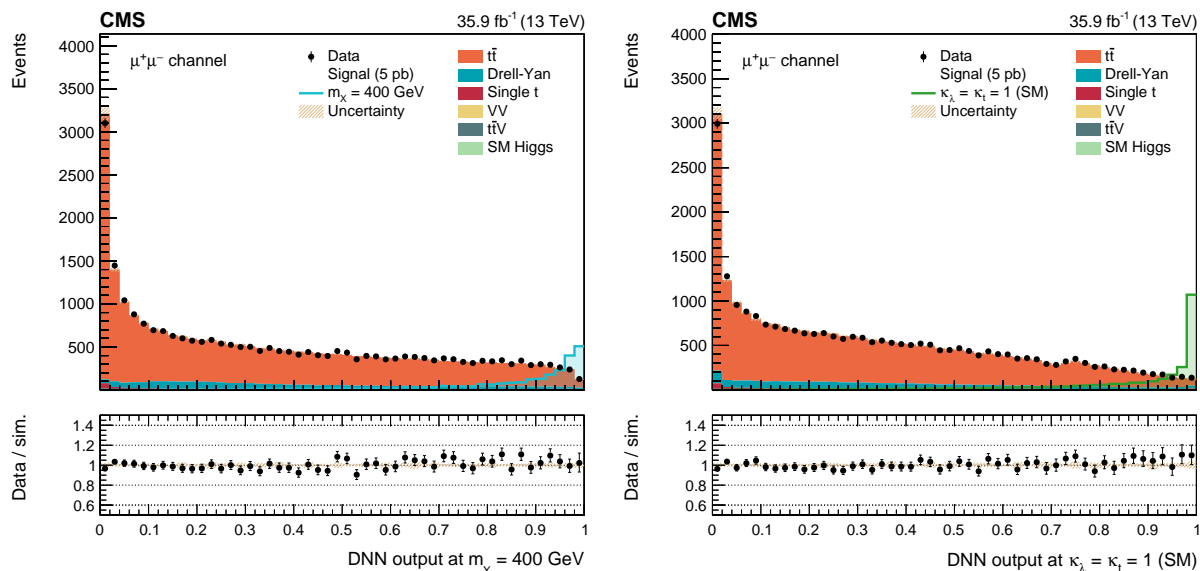


Figura 6: (Sinistra) La distribuzione DNN ottenuta considerando dati misurati e simulati nell'analisi di [14]. I valori attorno a 0 hanno caratteristiche del fondo, quelli attorno ad 1 sono di tipo segnale. Il risultato della DNN parametrica risonante è calcolato per $M_X = 400$ GeV.
 (Destra) Come a sinistra per DNN parametrica non risonante calcolata con $\kappa_\lambda = 1$ e $\kappa_t = 1$.
 (Left) The DNN output distribution in data and simulated events in the analysis of [14]. Output values towards 0 are background-like, while output values towards 1 are signal-like. The parameterized resonant DNN output is evaluated at $M_X = 400$ GeV.
 (Right) As in the left panel for the parameterized nonresonant DNN with output evaluated at $\kappa_\lambda = 1$ and $\kappa_t = 1$.



- [1] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (2012) 1 [arXiv:1207.7214].
- [2] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 (2012) 30 [arXiv:1207.7235].
- [3] [9] CMS collaboration, *Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV*, JHEP 06 (2013) 081 [arXiv:1303.4571].
- [4] <https://wlcg-public.web.cern.ch/about>
- [5] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 (2008) S08003.
- [6] CMS collaboration, *The CMS experiment at the CERN LHC*, 2008 JINST 3 S08004.
- [7] J. Neyman, E. Pearson, *On the problem of the most efficient tests of statistical hypotheses*. Phil. Trans. R. Soc. Lond. A 231, 694–706 (1933).
- [8] P. Baldi, p. Sadowski, D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*. Nat Commun 5, 4308 (2014), <https://doi.org/10.1038/ncomms5308>
- [9] The ATLAS Collaboration, *Search for heavy diboson resonances in semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, arXiv:2004.14636v1 [hep-ex] 30 Apr 2020, CERN-EP-2020-049, To appear EPJC.
- [10] A. Sherstinsky, *Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network*, Physica D 404 (2018) 132306, arXiv: 1808.03314.
- [11] F. Chollet et al., *Keras, GitHub repository (2015)*, <https://github.com/fchollet/keras>.
- [12] R. Al-Rfou et al. (Theano Development Team), *Theano: A Python framework for fast computation of mathematical expressions*, arXiv: 1605.02688 (2016).
- [13] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, D. Whiteson, *Parameterized neural networks for high-energy physics*, Eur. Phys. J. C 76 (2016) 235 [arXiv:1601.07913].
- [14] The CMS collaboration, *Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}l\nu l\nu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP01(2018)054.

Konstantinos Bachas: è un fisico, esperto in analisi dati, dell'Università Aristotele University di Salonicco, dove si è laureato in Fisica nel 2002. Ha ottenuto il Master of Science in fisica delle particelle elementari presso l'Università di Manchester (2003), e il dottorato di ricerca all'Università Aristotele nel 2008. È stato ricercatore presso il CERN (2009-2012), Salonicco (2013-2015), e la sezione INFN di Lecce (2016-2019). La sua attività di ricerca si è svolta principalmente nell'ambito delle attività dell'esperimento ATLAS, e ha riguardato diversi argomenti, dalla misura di processi di SM e ricerca di nuovi fenomeni nei dati di ATLAS alla del rivelatore ATLAS, con speciale attenzione allo spettrometro di muoni, e la messa a punto e verifica di sub-detector di ATLAS (rivelatori a *strip* del tracciatore interno a i tubi a *drift* dello spettrometro di muoni). Nel periodo passato a Lecce presso la sede dell'INFN è diventato esperto di tecniche moderne di ML, come DNN e RNN, che ha utilizzato per la ricerca della produzione non risonante di di-bosoni nell'analisi dei dati del run-2 di ATLAS. Il suo lavoro ha contribuito a consolidare l'uso di questi strumenti di ML come pratica comune nell'analisi dei dati di ATLAS.

Stefania Spagnolo: è professore associato di Fisica Nucleare e Subnucleare presso l'Università del Salento. È un componente della collaborazione ATLAS al CERN dal 2001 e ha recentemente contribuito all'esperimento PADME ai Laboratori Nazionali di Frascati (LNF) dell'INFN. Si è laureata e ha conseguito il Ph.D. in Fisica a Lecce, lavorando all'esperimento KLOE ai LNF. Dal 1998 al 2000 è stata ricercatrice del Rutherford Appletin Laboratory (UK) e componente della collaborazione OPAL a LEP (CERN). I suoi campi di interesse principali sono la fisica elettrodebole di precisione nel settore dei multibosoni e la ricerca di deviazioni dal MS che suggeriscano nuova fisica, compresi gli scenari di materia oscura non-WIMP.

Konstantinos Bachas: is a physicist and data scientist at the Aristotle University of Thessaloniki, where he graduated in Physics in 2002. He obtained a Master of Science in experimental particle physics at the University of Manchester (2003), and the Ph.D. at the Aristotle University in 2008. He was a research fellow at CERN (2009-2012), Thessaloniki (2013-2015), and INFN Lecce (2016-2019). His research activity has developed mainly within the ATLAS experiment, touching several areas, from measurements of Standard Model processes and searches for new phenomena with the ATLAS data to simulation of the ATLAS detector, with focus on the muon spectrometer, and ATLAS sub-detector commissioning and testing (strip detectors of the inner tracker and drift tubes of the muon spectrometer). During his contract with INFN, he became a crucial expert of modern Machine Learning techniques, like DNN and RNN, that he ported to the search for resonant di-boson production with the ATLAS data of run-2. His work contributed to establishing the use of these ML tools in the common practice of ATLAS data analysis.

Stefania Spagnolo: is professor of Nuclear and Subnuclear Physics at the University of Salento. She is working in the ATLAS experiment at CERN since 2001 and she joined in 2015 the PADME experiment at the Laboratori Nazionali di Frascati of INFN. She graduated and obtained the Ph.D. at the Università del Salento, working in the KLOE experiment at LNF. She was a research fellow at Rutherford Appleton Laboratory (UK) from 1998 to 2000 in the OPAL experiment at LEP (CERN). Her main fields of interest are tests of electroweak physics and search for deviations hinting to new phenomena connected to gauge and Higgs bosons, including scenarios for non-WIMP dark matter.