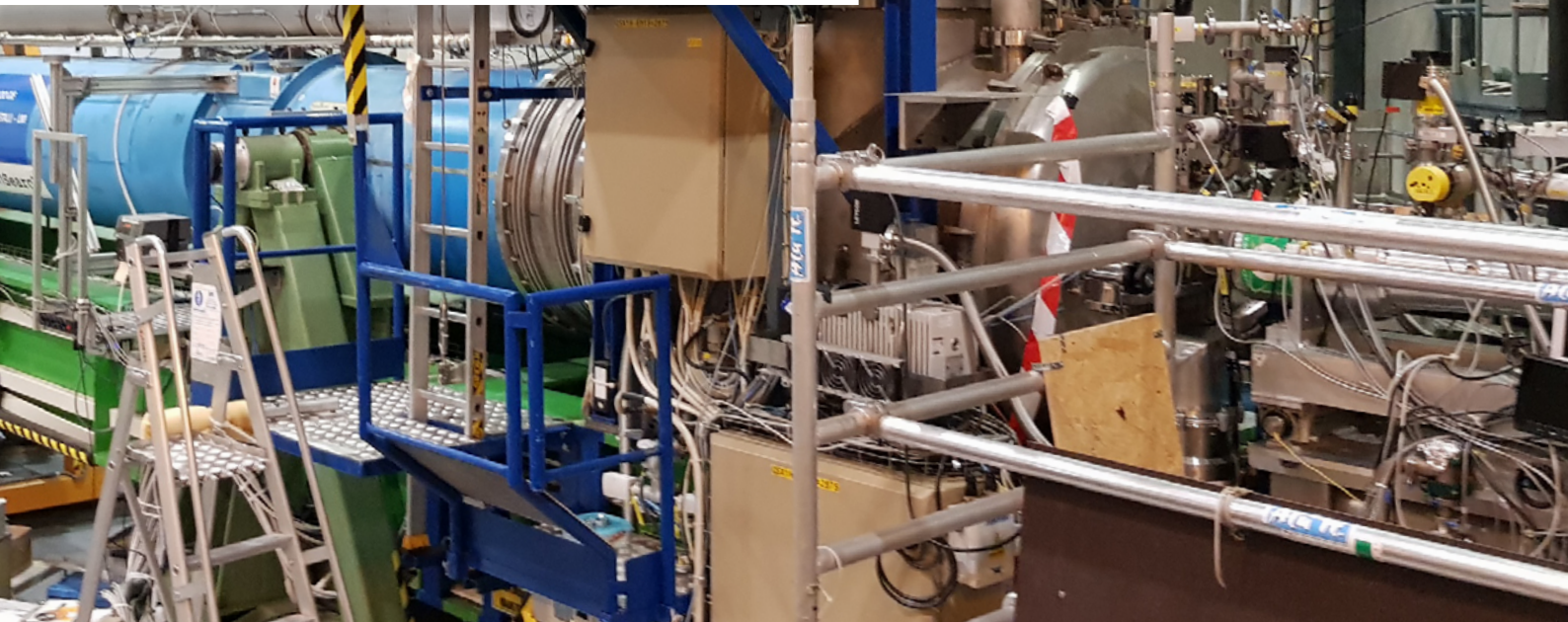


JungforscherInnen publizieren
online | *peer reviewed* | original

Verlag:
Physikalisch-
Technische
Bundesanstalt



Physik

Vorsortierer für Datenflut

Neuronale Netze auf der Suche
nach dunkler Materie

Die Dunkle Materie des Universums besteht vermutlich aus neuartigen Elementarteilchen, den Axionen. Diese sucht das CAST-Experiment am CERN in Genf. Um Signal- von Untergründereignissen zu trennen, wird das Potential eines Multilayer-Perceptrons, ein bestimmter Typ eines Neuronalen Netzwerks, als Klassifikationsmethode verwendet. Dies erweist sich als vielversprechender Ansatz.

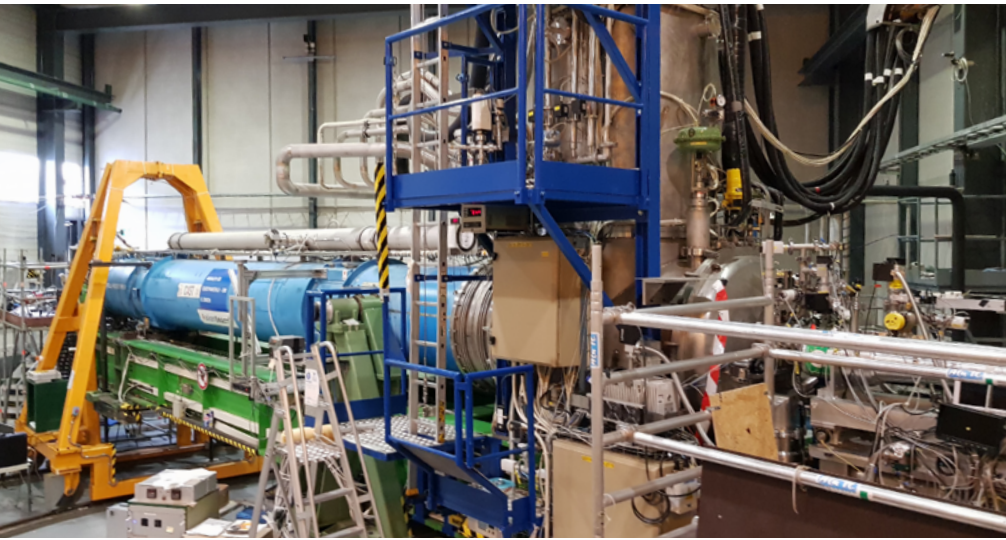
DIE JUNGFORSCHERIN



Carolin Kohl (2001)
Paul-Klee-Gymnasium
Overath

Eingang der Arbeit:
19.8.2019

Arbeit angenommen:
16.12.2019



Vorsortierer für Datenflut

Neuronale Netze auf der Suche nach dunkler Materie

1. Einleitung

Obwohl die dunkle Materie über 25 Prozent der gesamten Energie und Materie des Universums ausmacht, stellt sie die Physik trotz intensiver Forschungsarbeit bislang vor große Rätsel. Die uns ausmachende baryonische Materie hingegen stellt nicht einmal 5 Prozent der Materie- und Energiebilanz des Kosmos dar [1]. Deshalb sucht das CAST-Experiment am CERN in Genf nach möglichen Kandidaten für dunkle Materie und will diese indirekt in Anwesenheit eines starken Magnetfeldes über deren Umwandlung in Röntgenphotonen detektieren. So soll eine fundamentale physikalische Frage geklärt werden: Welche Materieform dominiert im Universum?

Doch auch zum Verständnis der baryonischen Materie kann das CAST-Experiment einen Beitrag leisten: Die Entdeckung eines bisher nur theoretisch eingeführten Teilchens – das sogenannte Axion – könnte das CP-Problems (Erhaltung der Charge-Parity-Invarianz) der starken Wechselwirkung lösen.

1.1 Das CAST-Experiment

Mithilfe des CAST-Experiments (CERN *Axion Solar Telescope*) sollen Axionen, die vermutlich in der Sonne entstehen und schließlich die Erde erreichen, durch deren Umwandlung in Röntgenphotonen nachgewiesen werden. Das theoretische Fundament dieser indirekten Detektionsmethode mittels Rönt-

genphotonen bildet der Sikivie-Effekt. Dieser Effekt sagt die Entstehung von Axionen durch Wechselwirkung thermaler Photonen mit den elektromagnetischen Feldern im Plasma des Sonneninneren sowie eine (Rück-)Umwandlung dieser Axionen in Röntgenphotonen in Gegenwart eines starken Magnetfeldes voraus [2].

Als Ziel des CAST-Experiments soll in einem bestimmten Massebereich nach der Kopplungskonstanten von Axionen und Photonen, welche die Häufigkeit und Wahrscheinlichkeit der Axion-Photon-Umwandlungen bestimmen, gesucht werden. Hierzu soll in einem nicht gebogenen Prototyp eines Large-Hadron-Collider-Dipolmagneten die Axion-Photon-Umwandlung durch das angelegte magnetische Feld ermöglicht werden. Die im Falle einer Umwandlung eines Axions entstehende schwache Röntgenstrahlung wird anschließend durch ein Spiegelsystem auf den Detektor fokussiert (siehe Abb. 1).

Die Röntgenphotonen gelangen durch ein dünnes transparentes Fenster in das Innere des Detektors. Dieser lässt sich grob in zwei Volumina unterteilen, die beide mit einer Mischung aus Argon-Gas und einem geringen Anteil Isobutan gefüllt sind und durch ein feinmaschiges Metallgitter mit einem Lochdurchmesser von ungefähr $30 \mu\text{m}$ voneinander getrennt sind (siehe Abb. 2a.). Im oberen Bereich – „*conversion gap*“ genannt – sollen über den Photoeffekt mittels der eintreffenden Photonen Elektronen aus den Gasatomen herausgeschlagen werden. Durch die anliegende Hochspannung werden die so ausgelösten Elektronen zum darunterliegenden Gitter gezogen, welches sie passieren und so in den zweiten Bereich des Detektors – die sogenannte „*amplification gap*“ – eintreten (siehe Abb. 2b). Durch die angelegte Spannung lösen die eintretenden Elektronen lawinenartig weitere Elektronen der umgebenden Gasatome aus. Alle ausgelösten Elektronen treffen

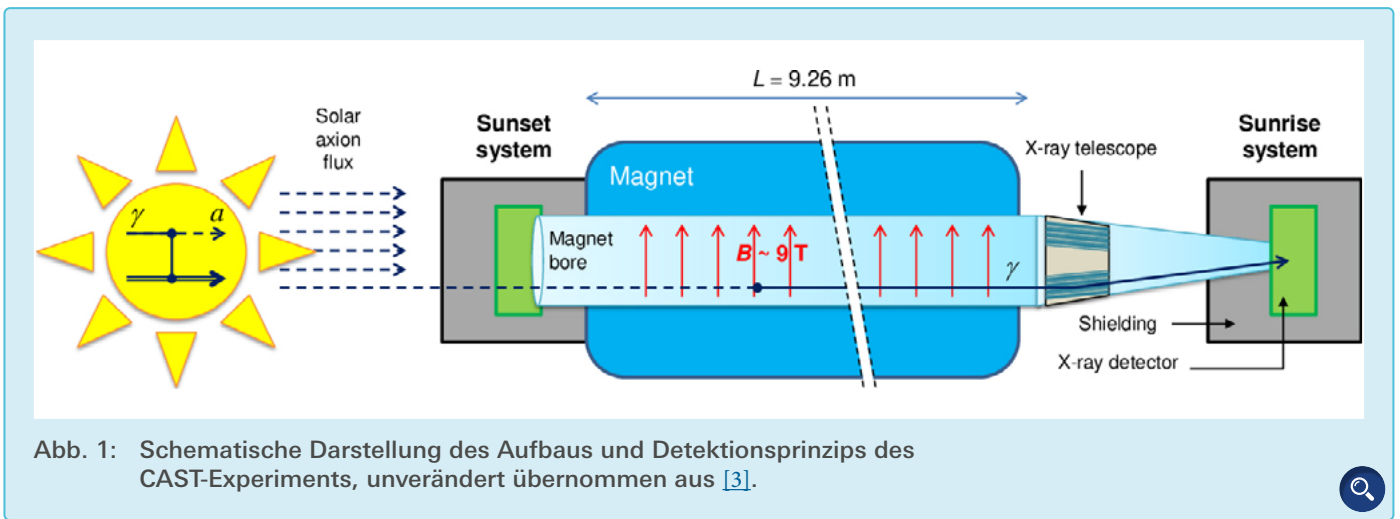


Abb. 1: Schematische Darstellung des Aufbaus und Detektionsprinzips des CAST-Experiments, unverändert übernommen aus [3].

nun – getrieben durch die Hochspannung – auf den Auslesebereich, wo sie durch pixelbasierte Chips detektiert werden.

Bei einem genau senkrecht auf den Detektor treffenden Teilchen ist eine sphärische Verteilung der detektierten Ladungsträger zu erwarten (Abb. 3a). Hierbei handelt es sich um potenzielle Signalereignisse, da diese durch Photonen aus der zu erforschenden Axion-Photon-Umwandlung verursacht werden können. Erhält man hingegen eine Art Spur (Abb. 3b), handelt es sich nicht um eines der gesuchten Photonen. Somit sind letztere Ereignisse als für das Experiment weniger interes-

sante Hintergrundereignisse herauszufiltern. Ursache für diese Hintergrundereignisse kann beispielsweise das, trotz vorhandener Bleiabschirmung nicht vollständig auszuschließende, Eindringen kosmischer Strahlung sein. Teilchen der kosmischen Strahlung, die zufällig genau senkrecht auf den Detektor treffen, lassen sich ggf. nicht von einem Röntgenphoton unterscheiden, da sie ebenfalls eine sphärische Ladungsverteilung auf dem Auslesechip verursachen.

Aus der detektierten Ladungsmenge und deren räumlicher Verteilung auf dem Chip werden die in Tab. 1 genannten Größen für jedes Ereignis errechnet [4].

Die Unterscheidung der aufgenommenen Ereignisse nach Signal und Untergrund erfolgt bereits mit Neuronalen Netzwerken. Für das Training dieser wurden einerseits Messungen des Hintergrundes verwendet, die aufgenommen wurden, als das Teleskop nicht auf die Sonne gerichtet war. Andererseits wurden für die Trainingsdaten der Signalereignisse Messdaten einer künstlichen Röntgenquelle benutzt, die in das Experiment eingeführt wurde.

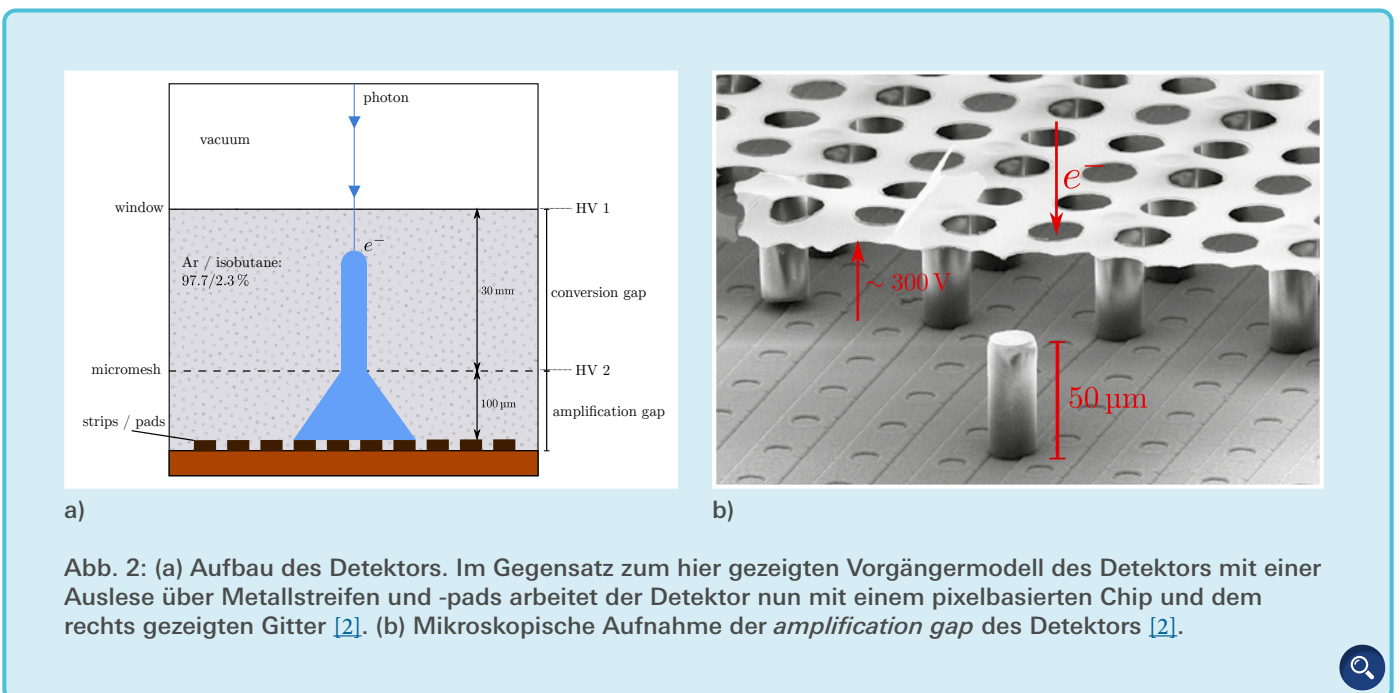


Abb. 2: (a) Aufbau des Detektors. Im Gegensatz zum hier gezeigten Vorgängermodell des Detektors mit einer Auslese über Metallstreifen und -pads arbeitet der Detektor nun mit einem pixelbasierten Chip und dem rechts gezeigten Gitter [2]. (b) Mikroskopische Aufnahme der *amplification gap* des Detektors [2].

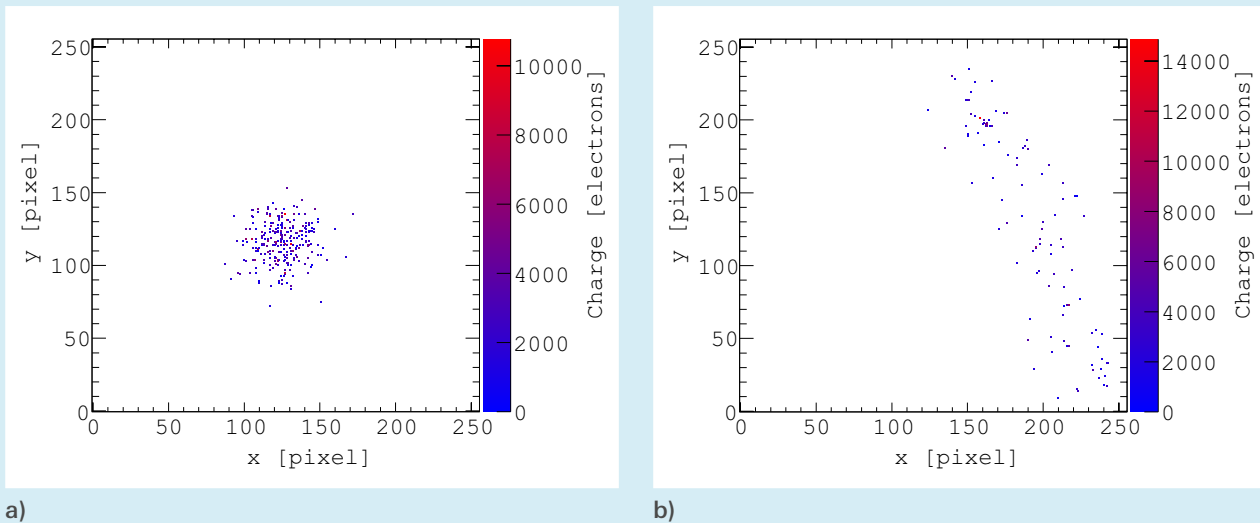


Abb. 3: Beispielhafte Ladungsverteilung auf dem Chip für Signal- (a) und Hintergrundereignisse (b)[5].



1.2 Fragestellung

Ziel des Projektes ist es, das Potential eines Multilayer-Perceptrons (MLP), eines Typ Neuronalen Netzwerks, abzuschätzen, wenn es als Klassifikationsmethode für Signal- und Untergrundereignisse eingesetzt wird.

In einer bereits bestehenden Arbeit [4] wurde ein Vergleich verschiedener Netzwerktypen – MLPs eingeschlossen – zur Datenauswertung vorgenommen. Im Unterschied zu dieser bereits existierenden Arbeit sollen nun die Optimierungsmöglichkeiten eines MLPs zur Datenklassifikation ausgelotet werden. Die zwischenzeitlich erfolgten Änderungen des Detektoraufbaus, der Datenauslese und die Neuberechnung der für das Training des Netzwerks verwendeten Parameter aus den ausgelesenen Daten sind weitere Gründe für die erneute Betrachtung dieser Methode zur Datenklassifikation. Folglich stellt die Abschätzung der zu erreichenden Signaleffizienz, also dem Verhältnis von Signalempfindlichkeit zur Untergrundunterdrückung, die Fragestellung der Arbeit dar. Anhand dieses Verhältnisses ist zu entscheiden, ob dieser Ansatz zur Datenauswertung weiterverfolgt werden kann und sollte.

2. Bisherige Datenklassifikation

Zum Zeitpunkt der Arbeit existierten bereits einige Ausarbeitungen in der CAST-Kollaboration zur Datenklassifikation mittels verschiedener Typen Neuronaler Netze, z. B. „*Boosted Decision Trees*“, „*Support-Vector-Machines*“ und auch zwei Varianten von MLPs. Eines der bereits existierenden MLPs, welches ebenfalls mittels *Backpropagation* trainiert wurde, erreichte u. a. eine Signaleffizienz von 97,49 Prozent bei einer Untergrundunterdrückung von 97,04 Prozent. Die entsprechende Accuracy betrug ca. 97,31 Prozent.

Dieses MLP verfügte über 20 Neuronen auf dem Input- und 25 Neuronen auf dem *Hidden-Layer*. Die verwendete *Learning-Rate* betrug 0,05. 30 Prozent der verfügbaren Daten wurden als Trainingsdatensatz verwendet. Die verwendeten Daten stammten hierbei aus Messungen mit einem Detektorprototyp unabhängig vom Versuchsaufbau am CERN. Mangels künstlicher Röntgenquellen, mit denen Ereignisse mit Photonen einer besonders niedrigen Energie erzeugt werden können, wurden diese Ereignisse aus gemessenen Ereignissen höherer Energie über den pro-

portionalen Zusammenhang von Energie und Anzahl detektierter Ladungen abgeleitet [4].

Da hierbei nicht mit den echten Daten des Experimentes, sondern mit im Labor erzeugten Datensätzen gearbeitet wurde und sich Detektoraufbau und -auslese ebenfalls verändert haben, knüpft die vorliegende Arbeit insofern an die bisherigen Ergebnisse an, dass nun die mit dem tatsächlichen Versuchsaufbau und der weiterentwickelten Datenauslese gemessenen Daten verwendet werden [4].

3. Methodik

3.1 Verwendetes Netzwerk

Das hier betrachtete Netzwerk ist in der Programmiersprache Python geschrieben worden. Diese Wahl ergab sich aus der vergleichbar einfachen und intuitiven Syntax und durch die weite Verbreitung dieser Sprache besonders im Bereich des *Machine Learning*, was Python besonders durch leicht verfügbare Literatur, Dokumentationen und vielfältige, sehr gut ausgebaute Bibliotheken besonders attraktiv macht. Für die Programmierung wurden folgende Bibliotheken verwendet: *numpy version*

1.18.1 [6], torch version 1.4.0 + cpu [7], sklearn version 0.22.2 post1 [8].

Die Wahl des *Multilayer-Perceptrons* (MLP) anstelle anderer Netzwerktypen liegt in der hohen Flexibilität dieser Netzwerkstruktur begründet, sodass eine große Wahrscheinlichkeit bestand, diesen Typ Netzwerk auch auf die gegebenen Messdaten des CAST-Experiments sinnvoll anwenden zu können. Darüber hinaus gestaltet sich die Programmierung dieses Netzwerktyps vergleichsweise einfach. So bot die Wahl eines in Python geschriebenen MLPs einen angemessenen Kompromiss aus komfortabler Handhabung und Anwendbarkeit auf das gegebene Problem. Als Trainingsmethode wurde *Backpropagation* gewählt.

Als anfängliche Vorlage diente ein MLP zur Klassifizierung des MNIST-Datensatzes. Dieser enthält handgeschriebene Ziffern und wird gerne zum Trainieren und Testen im Zusammenhang mit Machine-Learning verwendet [7]. MNIST ist hierbei eine Abkürzung für *Modified National Institute of Standards and Technology*.

Aufgrund der für dieses Projekt leicht zugänglichen Datenmengen wurde eine Unterteilung in einen Trainingsdatensatz und einen Testdatensatz mit dem Verhältnis 8:2 vorgenommen. Nach jeder Trainingsepoche wurde jeweils zur Evaluierung der Testdatensatz im Testmodus, d. h. ohne *Backpropagation* und Anpassung der erlernbaren Netzwerkparameter, durchlaufen. Nach Abschluss des gesamten Trainingsprozesses erfolgte abschließend eine Klassifikation des Testdatensatzes.

Die Test- und Trainingsdaten für Signal- und Hintergrundereignisse kamen aus derselben Messreihe, sodass Test- und Trainingsereignisse nicht anhand messreihenspezifischer Unterschiede erkannt werden können. Die Einteilung in Test- und Trainingsdatensatz erfolgte mittels eines *pseudo-random number generators* [9]. Um die Vergleichbarkeit

Tab. 1: Kurzerläuterung der errechneten Parameter [4].

Name des Parameters	Kurzerläuterung
eventNumber	Nummer zur Identifizierung des Ereignisses
hits	Anzahl der eingetroffenen Elektronen: $\text{Elektronen} = \frac{\text{Gesamtladung}}{\text{Elementarladung des Elektrons}}$
centerX, centerY	x- bzw. y- Koordinaten des Mittelpunkts der Ladungsdetektionen
length, width	Länge bzw. Breite der Spur auf dem Chip
eccentricity	Maßzahl für die Ellipsenförmigkeit, Verformung relativ zum Kreis
KurtosisLongitudinal	longitudinale bzw. transversale Kurtosis
kurtosisTransverse	Kurtosis: Maßzahl für die Stauchung in x-Richtung einer Verteilung im Vergleich zur Normalverteilung
rmsTransverse	Root-Mean-Square (RMS) des Abstandes der getroffenen Pixel von der Haupt- bzw. kleinen Halbachse der Spur
rmsLongitudinal'	RMS: Quadratwurzel des Durchschnitts der quadrierten Abweichung vom Mittelwert einer Verteilung (skalarer Wert)
fractionInTransverseRms	Anteil derjenigen angesprochenen Pixel mit einem Abstand kleiner oder gleich des als rmsTransverse errechneten Root-Mean-Square des Abstands der detektierten Ladungen zur kleinen Halbachse der Ladungsverteilung
skewnessLongitudinal	Asymmetrie der Verteilung der Ladungsträger in transversaler bzw. longitudinaler Richtung im Vergleich zur Normalverteilung
skewnessTransverse	
rotationAngle	Winkel zwischen der Hauptachse der elliptischen Verteilung und den zur Lokalisation der Pixel verwendeten Koordinatenachsen.

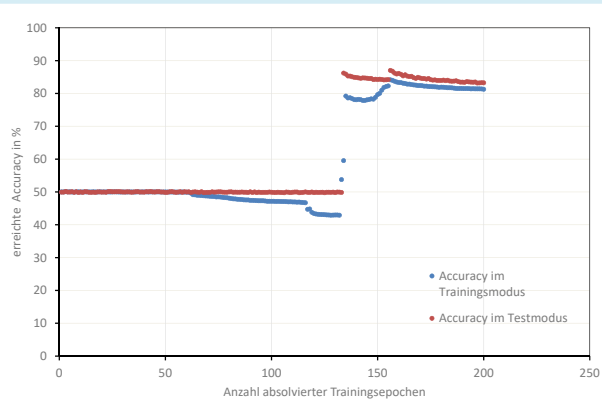
der Ergebnisse untereinander sicherzustellen, wurde der *pseudo-random number generator* mit unverändertem *Seed* verwendet. Für die Label wurde *One-Hot-Encoding* mit je einer Kategorie für Signal- und Hintergrundereignisse verwendet [10].

Die Anzahl der Neuronen auf dem *Input-Layer* entsprach der Anzahl der Parameter der Inputdaten (siehe Tab. 1), die Anzahl der Neuronen auf dem *Output-Layer* betrug 2, sodass erwartungsgemäß ein Neuron bei Signalereignissen und das andere bei Hintergrundereignissen eine höhere Aktivierung aufwies.

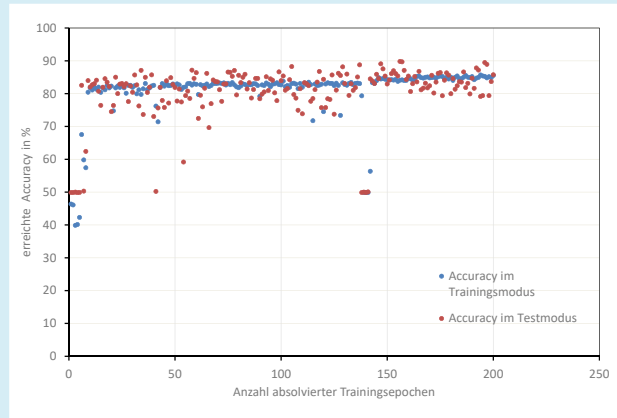
Nach jeder Veränderung der Netzwerkparameter erfolgte die Entscheidung über die nächste Variation anhand der erzeugten Grafiken.

3.2 Erzeugte Grafiken zur Analyse des Netzwerks

In einem Histogramm wurde die Aktivierung eines der beiden Neuronen auf dem Output-Layer bei der Klassifikation des Testdatensatzes nach Abschluss des Trainings aufgetragen. Je klarer die Trennung der verschiedenen Akti-



a)



b)

Abb. 4: Accuracy der Klassifizierung des Testdatensatzes und des Trainingsdatensatzes. Die abgebildeten Netzwerke besaßen je 2 Hidden-Layer mit jeweils 40 Neuronen und wurden über 200 Epochen mit einer Learning-Rate von 10^{-5} trainiert. Als Aktivierungsfunktion wurde die obig erwähnte Sigmoid-Funktion (a) bzw. $\tanh(x)$ (b) verwendet. Beide Netzwerke verwendeten Mean-Squared Error als Loss-Funktion und Mini-Batch Stochastic Gradient Descent mit einem Moment von 0,9. Ein Mini-Batch umfasste hierbei 100 Ereignisse.



vierungen, desto eindeutiger auch die Klassifizierung. Ergänzend wurde die Aktivierung des Signalneurons gegen die Aktivierung des Hintergrundneurons aufgetragen. Als Maß für die Leistungsfähigkeit des Netzwerks wird hier die Accuracy verwendet. Die Accuracy wird wie folgt berechnet: Formel 1.

Die über die Trainingsepoche gemittelte Accuracy wurde jeweils für den Trainings- und Testdatensatz in der Accuracy-Kurve in Abhängigkeit der Trainingsdauer, hier dargestellt in Trainingsepochen, aufgezeichnet. In ähnlicher Weise wurde mit dem über die Trainingsepoche gemittelten Loss verfahren. Auch dieser wurde in Abhängigkeit der Trainingsdauer, wieder in Einheiten von Trainingsepochen, veranschaulicht.

In der ROC-Kurve (Receiver Operating Characteristic curve) ist der pro-

zentuale Anteil der richtig als Signalereignis klassifizierten Ereignisse an der Gesamtzahl der Signalereignisse (True Positive Rate) als Funktion des prozentualen Anteils der fälschlicherweise als Signalereignis klassifizierten Ereignisse an der Gesamtzahl der Hintergrundereignisse (False Positive Rate) aufgetragen. Zur Berechnung der entsprechenden Wertepaare wird ein Schnittwert auf die Aktivierung des Signalneurons bei der Klassifikation des Testdatensatzes nach Abschluss des Trainings gesetzt, sodass dessen Aktivierung mindestens dem des Schnittwertes entsprechen muss, damit das Ereignis als Signalereignis klassifiziert wird. Für jeden Schnittwert ergibt sich dann eine True Positive Rate und die entsprechende False Positive Rate [11]. Des Weiteren wurde die True Negative Rate analog zum obigen Verfahren ermittelt und als Funktion der True Positive Rate dargestellt.

In Erwartung einer Steigerung der Leistungsfähigkeit des anfänglichen Netzwerks wurden verschiedene Netzwerkparameter [12] unabhängig voneinander variiert und der Effekt der Variation anhand der oben diskutierten Visualisierungsmethoden beurteilt, um über weitere sinnvolle Variationsmöglichkeiten zu entscheiden.

4. Optimierungsprozess

4.1 Variation der Aktivierungsfunktion

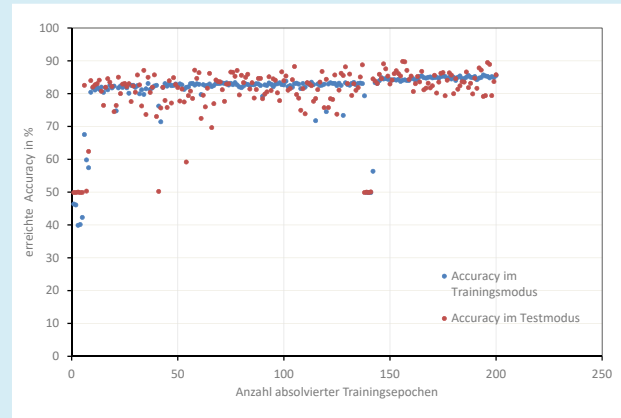
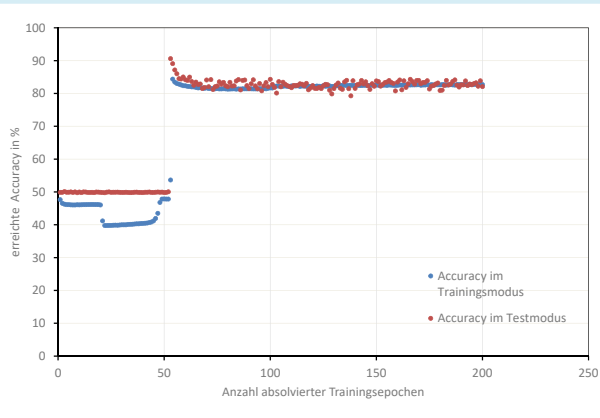
Bei Einsatz von $\tanh(x)$ anstelle einer Sigmoid-Funktion, definiert als

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}},$$

war eine deutliche Beschleunigung des Trainings zu bemerken. Da die zu Anfang verwendete Sigmoid-Funktion aufgrund der geringeren ersten Ablei-

$$\text{Accuracy} = \frac{\text{richtig klassifizierte Signalereignisse} + \text{richtig klassifizierte Hintergrundereignisse}}{\text{gesamte Anzahl der gemachten Klassifikationen}}$$

Formel 1



a)

b)

Abb. 5: Accuracy-Kurven bei Einsatz des *Nesterov Moments* (a) und des Adam-Trainingsalgorithmus (b)



nung zu kleineren Gradienten führte, scheint das Netzwerk mehr als doppelt so lange zum Erreichen der von beiden Netzwerken gezeigten maximalen *Accuracy* von rund 84 Prozent zu benötigen (siehe Abb. 4a). Das mit *tanh(x)* trainierte Netzwerk ließ sich durch eine Verkleinerung der Learning Rate auf 10^{-6} weiter verbessern, da so die Fluktuationen der *Accuracy* im Testdatensatz reduziert werden konnten (siehe Abb. 4b).

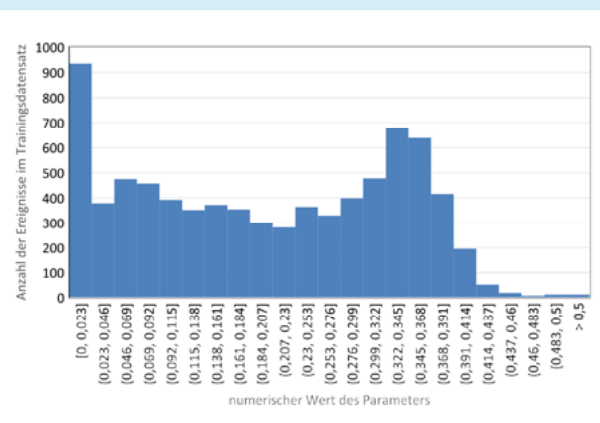
4.2 Variation des Trainingsalgorithmus

Um das Training weiter zu beschleunigen, wurden neben *Mini-Batch Stochastic Gradient Descent* (SGD) mit *Moment SGD* mit *Nesterov Moment* und *Adam* (*Adaptive Moment Estimation*) als Trainingsalgorithmen erprobt.

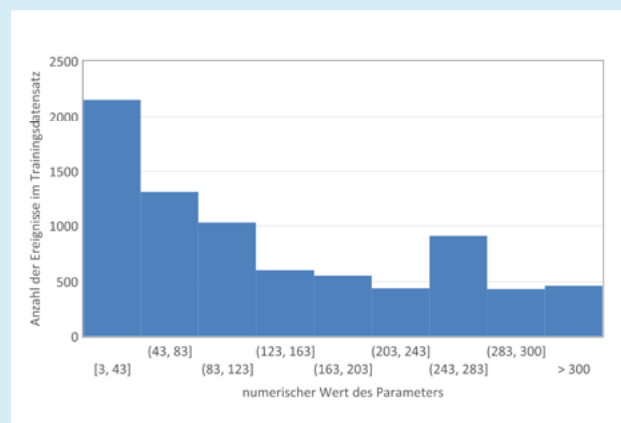
Ziel der Verwendung des *Nesterov Moments* nach Definition in [7] ist der Einbezug der über den Trainingsprozess bereits berechneten Gradienten. Hier-

durch mitteln sich starke Schwankungen in den Gradienten zu einem gewissen Grad heraus [13].

Durch den Einfluss des exponentiell gemittelten arithmetischen Mittels und der Quadrate der zuvor berechneten Gradienten kombiniert der Trainingsalgorithmus Adam RMSProp (exponentielles Mitteln der Quadrate der Gradienten) und Moment (exponentielles Mitteln der zuvor berechneten Gradienten). Hierdurch ist ebenfalls eine Beschleunigung des Lernprozesses zu



a)



b)

Abb 6: Die Verteilungen der Parameter der Inputdaten (*event shapes*) im Trainingsdatensatz weist eine sehr große Varianz auf. Hier ist exemplarisch die Verteilung des Inputparameter *fractionInTransverseRms* (a) und *hits* (b) gezeigt (siehe Tab 1 für Erläuterung der Parameter).



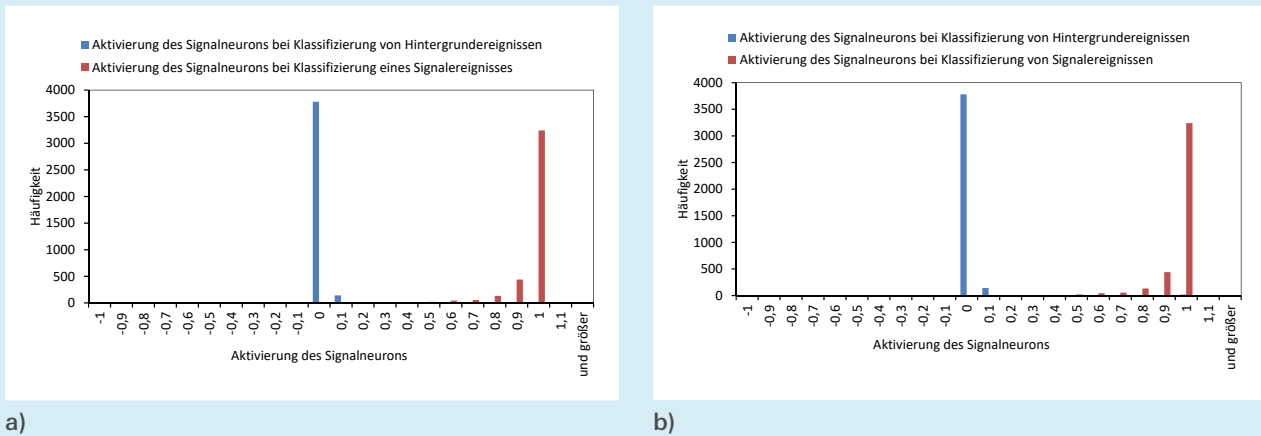


Abb. 7: (a) Aktivierung des Signalneurons bei Klassifizierung von Signal- und Hintergrundereignissen für das mit Normalisierung trainierte Netzwerk. (b) Aktivierung des Signalneurons bei Klassifizierung von Signal- und Hintergrundereignissen für das ohne Normalisierung trainierte Netzwerk.

erwarten. Für die Parameter β_1 und β_2 erweisen sich die Standardwerte von 0,9 bzw. 0,999 als gute Wahl. Die ersten und zweiten Momente der Verteilung der Gradienten im Iterationsschritt, hier als m_t bzw. v_t bezeichnet, werden in der in [14] vorgestellten Weise in jedem Trainingsschritt berechnet.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \cdot g_t$$

bzw. nach Bias-Korrektur

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$v_t = \beta_2 \cdot v_t - 1 + (1 - \beta_2) \cdot g_t^2$$

bzw. nach Bias-Korrektur

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

Wobei g_t für die in Iterationsschritt t berechneten Gradienten steht. Die Parameter im Iterationsschritt θ_t werden nach folgendem Prinzip in jedem Trainingsschritt berechnet:

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{\hat{m}_t}{(\sqrt{\hat{v}_t} + \epsilon)}$$

mit α als Learning Rate und ϵ als Konstante für numerische Stabilität.

Die Verwendung des *Nesterov Moments* resultierte lediglich in einer unwesentlichen Verbesserung des Netzwerks, die sich auf ca. 0,1 Prozent beschränkte. Bei Einsatz des Adam Trainingsalgorithmus war eine Erhöhung der *Accuracy* auf 96,8 Prozent in und rund 98,6 Prozent im Testdatensatz zu beobachten. Die leicht höhere *Accuracy* im Testdatensatz ist auf die im Vergleich zum Trainingsdatensatz geringere Datenmenge zurückzuführen, wodurch

der Testdatensatz eine kleinere Varianz aufweisen könnte. Zudem wurde durch Adam der Trainingsprozess beschleunigt. Die *Learning Rate* konnte für das mit Adam trainierte Netzwerk von 10^{-6} wieder auf 10^{-5} gesteigert werden.

4.3 Normalisierung

Da die Inputwerte eine große Spannweite von mehreren Größenordnungen abdeckten (siehe unten), erschien

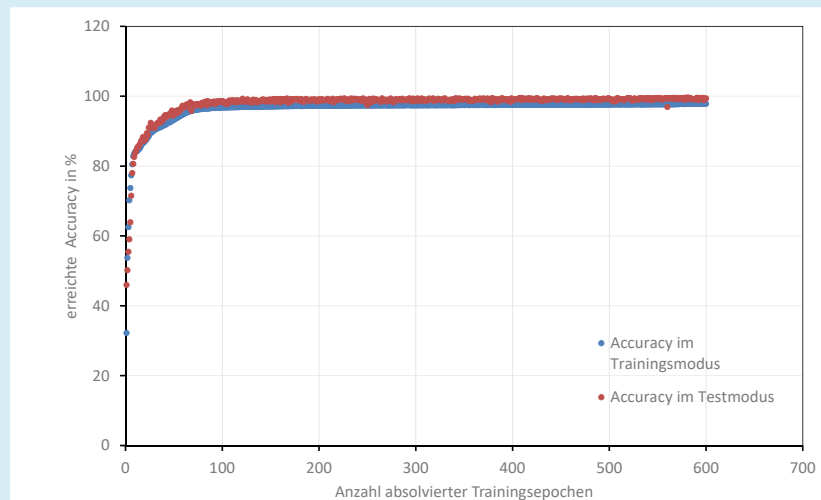
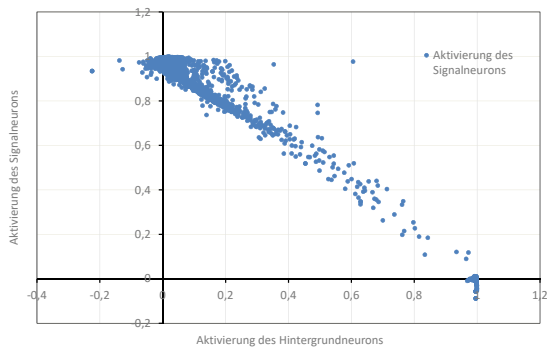
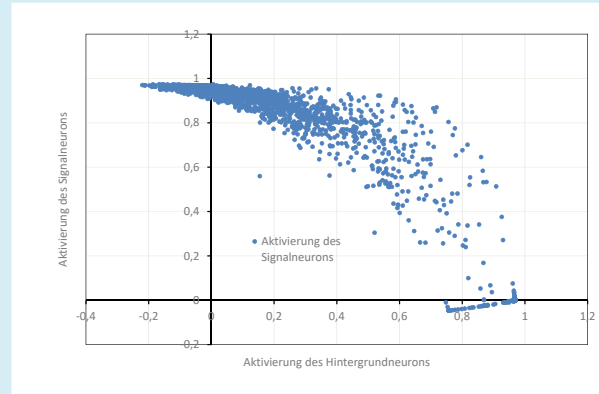


Abb. 8: *Accuracy* in Abhängigkeit der Trainingsdauer in Trainingsepochen. Das abgebildete Netzwerk besaß 3 *Hidden-Layer* zu je 69 Neuronen und wurde mit Adam als Trainingsalgorithmus über 600 Trainingsepochen trainiert. Als Aktivierungs- bzw. Loss-Funktion wurde $\tanh(x)$ bzw. MSE verwendet. Die *Learning Rate* betrug 10^{-5} .



a)



b)

Abb. 9: (a) Korrelation der Aktivierungen der Neuronen auf dem *Output-Layer* für das vergrößerte Netzwerk mit 3 *Hidden-Layern* zu je 69 Neuronen. (b) Korrelation der Aktivierungen der Neuronen auf dem *Output-Layer* für das Netzwerk mit 2 *Hidden-Layern* zu je 40 Neuronen.



die Normalisierung des Inputs sinnvoll. Ebenfalls wurde die Normalisierung der an die jeweils nächste Schicht von Neuronen übergebenen Aktivierungen, auch bekannt als *Batch Normalization*, erprobt.

So wurde zunächst eine Transformation vorgenommen, um eine Verteilung mit einer Varianz von 1 und einem Mittel von 0 zu halten.

Wider Erwarten resultierte die Normierung der Inputdaten in keiner merklichen Verbesserung, sodass beispielsweise die Verteilung der Aktivierungen des Signalneurons, wie in [Abb. 7](#) gezeigt, nur wenige Unterschiede aufwies.

4.4 Variation der Anzahl und Größe der *Hidden-Layer*

Die maximal im Testdatensatz erreichte *Accuracy* von ca. 99,6 Prozent im Trainingsdatensatz (siehe [Abb. 8](#)) konnte durch Erweiterung des Netzwerks von 2 auf 3 *Hidden-Layer* und der Erhöhung der Anzahl der Neuronen auf diesen *Hidden-Layern* auf jeweils 69 Neuronen noch gesteigert werden. Auch die *Accuracy* im Trainingsdatensatz ließ sich leicht auf ca. 97,8 Prozent erhöhen. Allerdings ist, wie in [Abb. 8](#) zu erkennen,

eine etwa dreimal so lange Trainingsdauer bis zum Erreichen eines Plateaus in der *Accuracy* nötig, da nun die Anzahl der zu erlernenden Parameter um die Anzahl der Gewichte und Bias-Werte der zusätzlichen Neuronen und *Hidden-Layer* erhöht ist. Die Aktivierungen der beiden Output-Neuronen sind gemäß [Abb. 9](#) korreliert.

5. Zusammenfassung und Fazit

Durch Variation der verschiedenen Parameter wie die *Learning Rate* und die Aktivierungsfunktion konnte die Leistungsfähigkeit des selbstgeschriebenen Netzwerks drastisch gesteigert werden. Zur Analyse des Einflusses der jeweiligen Parameter wurden selbstgenerierte Grafiken herangezogen. Mit der besten gefundenen Kombination der Parameter konnten schlussendlich ca. 99,6 Prozent aller Ereignisse richtig klassifiziert werden.

Die anfängliche Hypothese über die gute Eignung eines Neuronalen Netzwerks dieses Typs für die Datenklassifikation konnte somit im Rahmen dieser Arbeit bestätigt werden. Die erreichte Zuverlässigkeit der Klassifizierung macht diesen Typ Netzwerk, mehr noch

unter Berücksichtigung einer wahrscheinlichen Verbesserung durch Training mit mehr Inputdaten, zu einer vielversprechenden Alternative zur derzeit verwendeten Likelihood-Methode.

Danksagung

Diese Arbeit entstand im Rahmen der zweiwöchigen CERN-Projektwochen, eine Initiative des Netzwerks Teilchenwelt. Während dieser Zeit ermöglichte die Integration in die CAST-Kollaboration die Arbeit am Experiment vor Ort in Genf und die Erstellung dieser Arbeit.

An dieser Stelle möchte ich mich besonders bei Sebastian Schmidt bedanken, der mich während des gesamten Zeitraumes mit Rat und Tat unterstützt hat. Weiterhin gilt mein Dank Dr. Valeriani-Kaminski, Priv.-Doz. Dr. Philip Bechtle sowie Dr. Claudia Behne und Susanne Dührkoop, ohne die die Durchführung des Projekts am CERN nicht möglich gewesen wäre. Ebenso möchte ich Dr. Schiffbauer für seine Unterstützung bei organisatorischen Fragen und dem Gutachter für seine wertvollen Anregungen und Korrekturvorschläge zu diesem Artikel herzlich danken.

Literaturverzeichnis

- [1] Welt der Physik. [Online] Deutsche Physikalische Gesellschaft, Bundesministerium für Bildung und Forschung. [Zitat vom: 22.03.2020.] <https://www.weltderphysik.de/gebiet/uni-versum/dunkle-energie/>.
- [2] Schmidt, Sebastian. Search for particles beyond the SM using an InGrid detector at CAST. [Masterarbeit] s.l.: Universität Bonn, 2016.
- [3] Anastassopoulos, V., Aune, S., Barth, K. et al. New CAST limit on the axion–photon interaction. Nature Physics. Creative Commons License siehe: <http://creativecommons.org/licenses/by/4.0/>, 2017.
- [4] Schmidt, Sebastian. Verbesserung der Untergrundunterdrückung eines neuen CAST-Detektors mittels multivariater Methoden aus TMVA. [Bachelorarbeit] s.l.: Universität Bonn, 09 2013.
- [5] Krieger, Christoph. Search for solar chameleons with an InGrid based X-ray detector at the CAST experiment. Bonn: s.n., 2017.
- [6] NumPy. NumPy v1.18 Manual. [Online] The SciPy community, 05.02.2020 [Zitat vom: 20.03.2020] <https://numpy.org/doc/1.18/>.
- [7] PyTorch. [Online] Facebook, Inc, 2017. [Zitat vom: 20.03.2020] https://pytorch.org/tutorials/beginner/former_torchies/tensor_tutorial_old.html.
- [8] Pedregosa et. al. Scikit-learn: Machine Learning in Python., Journal of Machine Learning Research. 2011.
- [9] community, SciPy. Scipy.org. Random sampling (numpy.random). [Online] 26.07.2019 [Zitat vom: 20.03.2020] <https://docs.scipy.org/doc/numpy/reference/random/index.html>.
- [10] Brownlee, Dr. Jason. Machine Learning Mastery – Making Developers Awesome at Machine Learning. Why One-Hot Encode Data in Machine Learning? [Online] 28.07.2017 [Zitat vom: 20.03.2020] <https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/>.
- [11] Brownlee, Dr Jason. Machine Learning Mastery – Making Developers Awesome in Machine Learning. ROC Curves and Precision-Recall Curves for Imbalanced Classification. [Online] 14.01.2020 [Zitat vom: 20.03.2020] <https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-imbalanced-classification/>.
- [12] Raschka, Sebastian. Python Machine Learning. Birmingham, UK: Packt Publishing Ltd., 2015. ISBN: 978-1-78355-513-0.
- [13] Gylberth, Roan. Momentum Method and Nesterov Accelerated Gradient. [Online] Medium, 02.05.2018 [Zitat vom: 28.03.2020] <https://medium.com/konvergen/momentum-method-and-nesterov-accelerated-gradient-487ba776c987>.
- [14] Kingma, Diederik P. and Jimmy Ba. Adam: A Method for Stochastic Optimization. Cornell University. [Online] 2014. [Zitat vom: 21.03.2020] <https://arxiv.org/pdf/1412.6980.pdf>.
- [15] CERN- Accelerating Science. CAST. [Online] 20.03.2020. <https://home.cern/science/experiments/cast>.
- [16] Bin Shi, S.S. Iyengar. Mathematical Theories of Machine Learning - Theory and Applications. s.l.: Springer Nature Switzerland AG, 2020. ISBN 978-3-030-17075-2.
- [17] Liao, Yuan-Hong. GitHub. [Online] 13.03.2017. [Zitat vom: 20.03.2020] <https://github.com/andrewliao11/dni.pytorch/blob/master/mlp.py>.
- [18] (IAXO), The International Axion Observatory. ResearchGate. [Online] [Zitat vom: 20.03.2020] https://www.researchgate.net/figure/Basic-setup-of-an-axion-helioscope-converting-solar-axions-in-a-strong-laboratory_fig2_282535931.
- [19] Bonaccorso, Guiseppe. Mastering Machine Learning Algorithms: Eypert techniques to implement popular machine learning algorithms and fine-tune your models. Birmingham, UK: Packt Publishing Ltd., 2018. ISBN: 978-1-78862-111-3.

Publiziere auch Du hier!

Forschungsarbeiten von
Schüler/Inne/n und Student/Inn/en

In der Jungen Wissenschaft werden Forschungsarbeiten von SchülerInnen, die selbstständig, z. B. in einer Schule oder einem Schülerforschungszentrum, durchgeführt wurden, veröffentlicht. Die Arbeiten können auf Deutsch oder Englisch geschrieben sein.

Wer kann einreichen?

SchülerInnen, AbiturientInnen und Studierende ohne Abschluss, die nicht älter als 23 Jahre sind.

Was musst Du beim Einreichen beachten?

Lies die [Richtlinien für Beiträge](#). Sie enthalten Hinweise, wie Deine Arbeit aufgebaut sein soll, wie lang sie sein darf, wie die Bilder einzureichen sind und welche weiteren Informationen wir benötigen. Solltest Du Fragen haben, dann wende Dich gern schon vor dem Einreichen an die Chefredakteurin Sabine Walter.

Lade die [Erstveröffentlichungserklärung](#) herunter, drucke und fülle sie aus und unterschreibe sie.

Dann sende Deine Arbeit und die Erstveröffentlichungserklärung per Post an:

Chefredaktion Junge Wissenschaft

Dr.-Ing. Sabine Walter
Paul-Ducros-Straße 7
30952 Ronnenberg
Tel: 05109 / 561508
Mail: sabine.walter@verlag-jungewissenschaft.de

Wie geht es nach dem Einreichen weiter?

Die Chefredakteurin sucht einen geeigneten Fachgutachter, der die inhaltliche Richtigkeit der eingereichten Arbeit überprüft und eine Empfehlung ausspricht, ob sie veröffentlicht werden kann (Peer-Review-Verfahren). Das Gutachten wird den Euch, den AutorInnen zugeschickt und Du erhältst gegebenenfalls die Möglichkeit, Hinweise des Fachgutachters einzuarbeiten.

Die Erfahrung zeigt, dass Arbeiten, die z. B. im Rahmen eines Wettbewerbs wie **Jugend forscht** die Endrunde erreicht haben, die besten Chancen haben, dieses Peer-Review-Verfahren zu bestehen.

Schließlich kommt die Arbeit in die Redaktion, wird für das Layout vorbereitet und als Open-Access-Beitrag veröffentlicht.

Was ist Dein Benefit?

Deine Forschungsarbeit ist nun in einer Gutachterzeitschrift (Peer-Review-Journal) veröffentlicht worden, d. h. Du kannst die Veröffentlichung in Deine wissenschaftliche Literaturliste aufnehmen. Deine Arbeit erhält als Open-Access-Veröffentlichung einen DOI (Data Object Identifier) und kann von entsprechenden Suchmaschinen (z. B. BASE) gefunden werden.

Die Junge Wissenschaft wird zusätzlich in wissenschaftlichen Datenbanken gelistet, d. h. Deine Arbeit kann von Experten gefunden und sogar zitiert werden. Die Junge Wissenschaft wird Dich durch den Gesamtprozess des Erstellens einer wissenschaftlichen Arbeit begleiten – als gute Vorbereitung auf das, was Du im Studium benötigst.



Richtlinien für Beiträge

Für die meisten Autor/Inn/en ist dies die erste wissenschaftliche Veröffentlichung. Die Einhaltung der folgenden Richtlinien hilft allen – den Autor/innen/en und dem Redaktionsteam

Die Junge Wissenschaft veröffentlicht Originalbeiträge junger AutorInnen bis zum Alter von 23 Jahren.

- Die Beiträge können auf Deutsch oder Englisch verfasst sein und sollten nicht länger als 15 Seiten mit je 35 Zeilen sein. Hierbei sind Bilder, Grafiken und Tabellen mitgezählt. Anhänge werden nicht veröffentlicht. Deckblatt und Inhaltsverzeichnis zählen nicht mit.
- Formulieren Sie eine eingängige Überschrift, um bei der Leserschaft Interesse für Ihre Arbeit zu wecken, sowie eine wissenschaftliche Überschrift.
- Formulieren Sie eine kurze, leicht verständliche Zusammenfassung (maximal 400 Zeichen).
- Die Beiträge sollen in der üblichen Form gegliedert sein, d. h. Einleitung, Erläuterungen zur Durchführung der Arbeit sowie evtl. Überwindung von Schwierigkeiten, Ergebnisse, Schlussfolgerungen, Diskussion, Liste der zitierten Literatur. In der Einleitung sollte die Idee zu der Arbeit beschrieben und die Aufgabenstellung definiert werden. Außerdem sollte sie eine kurze Darstellung schon bekannter, ähnlicher Lösungsversuche enthalten (Stand der Literatur). Am Schluss des Beitrages kann ein Dank an Förderer der Arbeit, z. B. Lehrer und Sponsoren, mit vollständigem Namen angefügt werden. Für die Leser kann ein Glossar mit den wichtigsten Fachausdrücken hilfreich sein.
- Bitte reichen Sie alle Bilder, Grafiken und Tabellen nummeriert und zusätzlich als eigene Dateien ein. Bitte geben Sie bei nicht selbst erstellten Bildern, Tabellen, Zeichnungen, Grafiken etc. die genauen und korrekten Quellenangaben an (siehe auch [Erstveröffentlichungserklärung](#)). Senden Sie Ihre Bilder als Originaldateien oder mit einer Auflösung von mindestens 300 dpi bei einer Größe von 10 · 15 cm! Bei Grafiken, die mit Excel erstellt wurden, reichen Sie bitte ebenfalls die Originaldatei mit ein.
- Vermeiden Sie aufwendige und lange Zahlentabellen.
- Formelzeichen nach DIN, ggf. IUPAC oder IUPAP verwenden. Gleichungen sind stets als Größengleichungen zu schreiben.
- Die Literaturliste steht am Ende der Arbeit. Alle Stellen erhalten eine Nummer und werden in eckigen Klammern zitiert (Beispiel: Wie in [12] dargestellt ...). Fußnoten sieht das Layout nicht vor.
- Reichen Sie Ihren Beitrag sowohl in ausgedruckter Form als auch als PDF

ein. Für die weitere Bearbeitung und die Umsetzung in das Layout der Jungen Wissenschaft ist ein Word-Dokument mit möglichst wenig Formatierung erforderlich. (Sollte dies Schwierigkeiten bereiten, setzen Sie sich bitte mit uns in Verbindung, damit wir gemeinsam eine Lösung finden können.)

- Senden Sie mit dem Beitrag die [Erstveröffentlichungserklärung](#) ein. Diese beinhaltet im Wesentlichen, dass der Beitrag von dem/der angegebenen AutorIn stammt, keine Rechte Dritter verletzt werden und noch nicht an anderer Stelle veröffentlicht wurde (außer im Zusammenhang mit **Jugend forscht** oder einem vergleichbaren Wettbewerb). Ebenfalls ist zu versichern, dass alle von Ihnen verwendeten Bilder, Tabellen, Zeichnungen, Grafiken etc. von Ihnen veröffentlicht werden dürfen, also keine Rechte Dritter durch die Verwendung und Veröffentlichung verletzt werden. Entsprechendes [Formular](#) ist von der Homepage www.junge-wissenschaft.ptb.de herunterzuladen, auszudrucken, auszufüllen und dem gedruckten Beitrag unterschrieben beizulegen.
- Schließlich sind die genauen Anschriften der AutorInnen mit Telefonnummer und E-Mail-Adresse sowie Geburtsdaten und Fotografien (Auflösung 300 dpi bei einer Bildgröße von mindestens 10 · 15 cm) erforderlich.
- Neulingen im Publizieren werden als Vorbilder andere Publikationen, z. B. hier in der Jungen Wissenschaft, empfohlen.

Impressum

[JUNGE]
wissenschaft



Junge Wissenschaft

c/o Physikalisch-Technische
Bundesanstalt (PTB)
www.junge-wissenschaft.ptb.de

Redaktion

Dr. Sabine Walter, Chefredaktion
Junge Wissenschaft
Paul-Ducros-Str. 7
30952 Ronnenberg
E-Mail: sabine.walter@verlag-jungewissenschaft.de
Tel.: 05109 / 561 508

Verlag

Dr. Dr. Jens Simon,
Pressesprecher der PTB
Bundesallee 100
38116 Braunschweig
E-Mail: jens.simon@ptb.de
Tel.: 0531 / 592 3006
(Sekretariat der PTB-Pressestelle)

Design & Satz

Sabine Siems
Agentur „provieler werbung“
E-Mail: info@provieler-werbung.de
Tel.: 05307 / 939 3350

