

Entwicklung von Methoden zum Nachweis von ökologisch erzeugten Produkten am Beispiel der Lachszucht - Neronale Netzanalyse -

Development of Methods to Detect Products Made from Organic Salmon

FKZ: 02OE073/1

Projektnehmer:

Max Rubner-Institut
Bundesforschungsinstitut für Ernährung und Lebensmittel
Hermann-Weigmann-Straße 1, 24103 Kiel
Tel.: +49 431 609-1
Fax: +49 431 609 2222
E-Mail: poststelle@mri.bund.de
Internet: <http://www.mri.bund.de>

Autoren:

Meisel, Hans

Gefördert vom Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz
im Rahmen des Bundesprogramms Ökologischer Landbau (BÖL)

Projekt 02OE073/1 "Ökofina"

Neuronale Netzanalyse der Analysedaten zum Nachweis ökologisch erzeugten Lachses

Einleitung

Das Hauptziel des Projektes, nämlich ein Analyseverfahren zum Nachweis von ökologisch erzeugtem Lachs zu finden, wurde durch die Analytik zur Bestimmung der Astaxanthin-Isomere erreicht (s. Schlussbericht). Allerdings wird der Nutzen dieses Verfahrens dadurch eingeschränkt, dass konventioneller Farmlachs, der wie Ökolachs mit Pigmenten aus der Phaffia-Hefe gefüttert wurde, fälschlicherweise als Ökolachs identifiziert würde. Deshalb bestand ein wesentliches Ziel der neuronalen Netzanalyse in der Auffindung eines neuronalen Netzwerks, das eine Identifizierung weitgehend fütterungsunabhängig ermöglicht.

Die neuronale Netzanalyse ist ein besonders geeignetes chemometrisches Verfahren zur Merkmalsextraktion komplexer analytischer Daten. Der Begriff Chemometrie umschreibt allgemein computerunterstützte mathematische Verfahren, die zur Gewinnung maximaler Information bei der Analyse von chemischen Daten verwendet werden. Die Auswertung mittels Chemometrie bedeutet vor allem die Extraktion und Interpretation von Informationen aus Messergebnissen, aus denen diese Informationen nicht unmittelbar hervorgehen. Hierfür stehen verschiedene mathematische Verfahren zur Datenreduktion und -interpretation zur Verfügung, wie z.B. künstliche neuronale Netzwerke (Artificial Neural Networks, ANN).

Bei künstlichen neuronalen Netzwerken handelt es sich um informationsverarbeitende Systeme, die aus einer großen Zahl einfacher Schalteinheiten zusammengesetzt sind und biologischen Neuronen insofern ähneln, als in solchen technischen oder künstlichen Neuronen Information ebenfalls durch Aktivierung und Hemmung verarbeitet wird. Das einfachste neuronale Netz besteht aus einer Eingabeschicht (Input-Neuron) und einem einzelnen Knoten, der Ausgabeschicht bzw. dem Ausgabeneuron (Output-Neuron). Darüber hinaus können zwischen Eingabe- und Ausgabeschicht weitere Zwischenschichten (verdeckte bzw. verborgene Schichten; engl.: hidden layers) vorhanden sein. Die Verknüpfung der

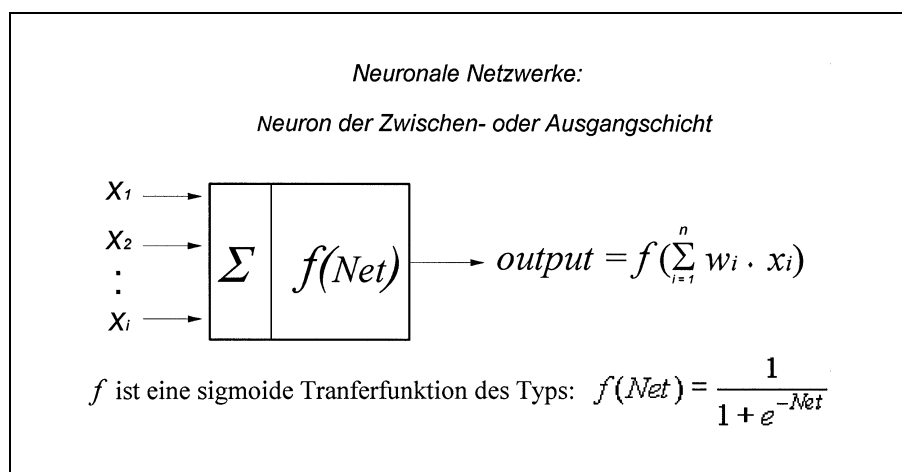


Abb. 1: Zweilagiges, vorwärtsverknüpftes neuronales Netzwerk
(x_i : Eingabewerte; w_i : Kantengewichte)

Neuronen erfolgt über sogenannte Kantengewichte (**Abb. 1**), d.h. die Neuronen einer Schicht sind über die Gewichtungen vollständig mit der vorherigen Schicht verbunden. Im Ausgabeneuron werden die von der Eingabeschicht erhaltenen Eingabewerte x_i mit den entsprechenden Kantengewichten w_i multipliziert

und aufsummiert. Unter Berücksichtigung eines Schwellenwertes wird der resultierende Wert durch eine (in der Regel nicht lineare) Transferfunktion in einen Ausgabewert umgerechnet und läuft als Signal zum folgenden Neuron (Zwischen- oder Ausgabeschicht) weiter. In simulierten neuronalen Netzwerken werden die Kantengewichte variabel eingegeben. Netzwerke werden trainiert, indem sich die Gewichtung ändert. Tatsächlich wird ein ANN durch eine Abfolge von Computer-gestützten Rechenoperationen verwirklicht.

Netzwerke, die neben den Eingangs- und Ausgangsschichten noch Zwischenschichten enthalten, sind besonders geeignet, komplexe, nicht lineare Input-Output-Beziehungen zu vollziehen. Das bekannteste Trainingsverfahren (Lernprozess) für Netzwerke mit drei oder mehr Schichten beruht auf dem Backpropagation-Algorithmus. Hierbei wird der Output-Fehler zunächst an die vor der Output-Schicht liegende Zwischenschicht zurückgemeldet und die Kantengewichte geändert. In einem dreischichtigen neuronalen Netzwerk werden also zunächst die Kantengewichte der Neuronen der Ausgangsschicht und dann die der Mittelschicht verändert. Da der jeweilige Fehler das Netzwerk rückwärts durchläuft, wird diese Lernregel als "Backpropagation of errors"-Regel (Backpropagation-Regel) bezeichnet.

Material und Methoden

Zur neuronalen Netzanalyse wurden folgende kommerziellen Software-Produkte verwendet: "NeuroSolutions v.4.32 (NeuroDimension, Inc., Gainesville, FL 32609, USA) in Kombination mit "NeuroSolutions for Excel" und der Tabellenkalkulation Excel (Microsoft Corp., Redmond, USA) sowie EasyNN-plus v.4.0g (Neural Planner Software, Stephen Wolstenholme, UK).

Für die chemometrische Analyse zur Klassifizierung von Ökolachs, Farmlachs und Wildlachs wurden verschiedene Kombinationen der kompositionellen Parameter in die Eingabeschicht (Input) des Netzwerkes eingegeben (**Abb. 2**). Zur Klassifizierung der unterschiedlich erzeugten Lachse wurden drei Ausgabe-Neuronen (Output) vom binären Typ verwendet. Der gesamte Datensatz von 100 Proben (s. Schlussbericht) wurde in drei Unter-Datensätze aufgeteilt. Der erste dieser Unter-Datensätze (70 %) wurde für das Training verwendet und ein zweiter Satz (10 %) für die Validierung des Netzwerkes zur Kontrolle des Trainingsprozesses. Die restlichen 20 % des Datensatzes wurden zum Testen des trainierten Netzwerkes eingesetzt (Testdatensatz), um zu prüfen, ob die nicht im Trainings- und Validierungsdurchlauf eingesetzten Lachsproben richtig vom trainierten Netzwerk zugeordnet werden. Das Training des Netzes erfolgte mit dem Backpropagation-Algorithmus über maximal 10000 Trainingszyklen. Bei der Klassifizierung ergeben Output-Werte von $> 0,5$ und $< 0,5$ die binäre Entscheidung 1 bzw. 0.

Zur Auffindung der besten Netzwerkkonfiguration wurden softwareseitig verschiedene Einstellungen zur Beeinflussung des Trainingsprozesses vorgenommen. Die Optimierung erfolgte durch multiples Training, um die Einstellungen zu finden, die einen minimalen Fehler ergeben. Nach entsprechenden Vorversuchen stellte sich heraus, dass mit einer Konfiguration mit einer Zwischenschicht, die aus vier bis sechs Neuronen besteht, die besten Klassifizierungsergebnisse erreicht werden können. Die Berechnung neuronaler Netzwerke mit verschiedenen Kombinationen von Analyseparametern als Eingabewerte wurden mit der Software NeuroSolutions mit folgenden Einstellungen vorgenommen: Zwischenschicht mit 4 Neuronen (optimiert durch "genetischen Algorithmus"), Transferfunktion = TanhAxon (hyperbolische Tangente $-1/+1$), Momentum = 0,7, Trainingszyklen = 10000.

Ergebnisse und Diskussion

In der **Tab. 1** sind die Ergebnisse für verschiedene Netzwerkkonfigurationen angegeben, die zur Identifizierung der Produktionsweise von Lachs eingesetzt wurden. Netzwerke mit verschiedenen Kombinationen von Analyseparametern als Eingabewerte, ergaben nach neuronaler Netzanalyse korrekte Resultate in Bezug auf den Testdatensatz, d.h. von Proben, die nicht für den Trainings- und

Validierungslauf verwendet worden waren, sowie eine 100 % korrekte Klassifikation aller Lachsproben des gesamten Datensatzes.

Es wird deutlich, dass bestimmte Kombinationen der Analysedaten (Eingabewerte) zu unterschiedlichen Bewertungen bzw. zu einer unterschiedlichen Anzahl richtiger Ergebnisse führen (**Tab. 1**). Zahlreiche Netzwerke haben die gewünschte Generalisierungsfähigkeit und sind in der Lage, die richtigen Ausgabewerte für die jeweiligen Eingangswerte zu liefern, d.h. die drei Produktionsweisen von Lachs korrekt zuzuordnen. Ein hoher Informationsgehalt für die Zuordnung zu einer der drei Produktionsweisen von Lachs ist in den Analysedaten der Stabilisotopen, insbesondere $d^{15}N$, sowie der Fettsäuren vorhanden. Mit der Eingabe der Stabilisotopenwerte in Kombination mit den Fettsäuren Linolen-, Eicosensäure und Arachidonsäure wird zu 100 % eine richtige Zuordnung erreicht (**Abb. 3**). Die Bedeutung dieser kompositionellen Parameter als wichtigste Einflussfaktoren auf die korrekte Klassifikation wird durch die Software-seitig implementierte Sensitivitätsanalyse und Analyse des Input-Einflusses bestätigt (**Abb. 4**).

Zusammenfassend kann festgestellt werden, dass die neuronale Netzanalyse bestimmter kompositioneller Parameter genügt, um eine Zuordnung der Produktionsweisen von Lachs bzw. den Nachweis ökologisch erzeugten Lachses zu erreichen. Ein entscheidender Punkt für die erfolgreiche Anwendung neuronaler Netze zur Klassifikation unbekannter Proben ist das Vorhandensein repräsentativer Daten und deren Aufteilung in einen Trainings-, Validierungs- und Testdatensatz. Die der hier gezeigten neuronalen Netzanalyse zugrunde liegenden Daten basieren auf 100 untersuchten Lachsen, die jedoch aus nur 8 verschiedenen Schwärmen stammen. In weiteren Versuchen sollte deshalb überprüft werden, ob weitere Proben, die nicht zu dem vorhandenen Datensatz gehören, eine 100 % korrekte Klassifikation der Produktionsweisen durch die berechneten neuronalen Netzwerke ergeben.

Input											Output				
Probe	d15N	d13C	Pentadecan			Octadecanoi			Oleic			Farmlachs	Ökolachs	Wildlachs	
			Myristic	oic	Palmitic	Palmitoleic	Stearic	c							
FL1.1	11.43	-21.96	7.48	0.54	14.61	10.61	2.03	14.35	6.17		1	0	0		
FL1.2	11.55	-22.55	7.16	0.52	14.57	10.03	2.2	14.47	5.75		1	0	0		
FL1.3	11.84	-22.06	8.73	0.62	14.04	10.68	2.19	13.48	6.83		1	0	0		
FL1.4	11.74	-21.71	7.5	0.55	14.03	10.39	2.13	13.71	6.5		1	0	0		
FL1.5	11.53	-21.77	8.14	0.6	15.09	11.11	2.26	13.94	6.75		1	0	0		
FL1.6	11.79	-22.07	8.52	0.61	14.68	10.69	2.18	13.49	6.45		1	0	0		
FL1.7	11.84	-21.66	7.97	0.6	13.5	9.63	2.16	12.49	6.05		1	0	0		
FL1.8	11.65	-21.74	7.2	0.57	16.03	11.58	2.35	14.25	6.41		1	0	0		
FL1.9	11.43	-21.61	7.11	0.56	14.24	10.93	2.12	13.43	6.48		1	0	0		
FL1.10	11.45	-21.96	7.5	0.58	13.62	10.61	2.05	13.86	5.85		1	0	0		
FL1.11	11.88	-21.54	6.71	0.54	15	10.24	2.31	13.97	5.94		1	0	0		
FL1.12	11.62	-21.67	7.43	0.59	15.82	11.29	2.28	14.35	6.36		1	0	0		
FL1.13	11.6	-21.51	7.63	0.59	14.65	11.29	2.03	13.14	6.35		1	0	0		
FL1.14	12.09	-21.6	7.73	0.58	13.77	10.86	2.12	13.89	5.87		1	0	0		
FL1.15	11.54	-22.1	7.93	0.59	13.06	10.89	2.08	13.58	6.09		1	0	0		
FL1.16	11.72	-22.48	7.65	0.59	13.64	10.87	2.02	13.67	6.07		1	0	0		
FL1.17	11.68	-22.39	8.42	0.63	13.47	11.75	1.95	13.99	6.19		1	0	0		
FL1.18	11.85	-21.81	8.15	0.62	13.93	10.91	2.01	13.4	6.2		1	0	0		
FL1.19	11.71	-21.57	7.64	0.6	13.66	10.64	2.03	13.54	6.31		1	0	0		
FL1.20	11.65	-22.21	8.74	0.65	14.33	11.2	1.96	13.65	6.3		1	0	0		
OL1.1	13.75	-19.68	8.52	0.77	16.05	9.83	2.83	14.48	6.36		0	1	0		
OL1.2	13.83	-19.35	8.97	0.83	16.82	10.42	2.74	14.37	6.31		0	1	0		
OL1.3	13.98	-20.44	9.24	0.85	17.03	10.28	2.64	14.83	6.3		0	1	0		
OL1.4	13.98	-20.23	9.03	0.84	16.56	10.41	2.74	14.94	6.29		0	1	0		
OL1.5	13.51	-20.1	8.31	0.79	16.78	9.96	2.94	14.12	6.47		0	1	0		
OL1.6	13.83	-19.98	9.58	0.87	17.02	10.6	2.76	14.97	6.29		0	1	0		
OL1.7	14.04	-19.51	9.25	0.82	15.28	10.56	3.72	13.36	6.67		0	1	0		
OL1.8	13.83	-19.94	8.21	0.75	16.57	9.76	2.87	14.86	6.07		0	1	0		
OL1.9	13.64	-20.45	8.09	0.74	16.35	9.44	2.55	14.58	5.85		0	1	0		
OL1.10	13.76	-20.16	8.28	0.76	15.66	9.48	2.9	14.51	6.19		0	1	0		
OL1.11	13.84	-19.07	7.92	0.7	16.42	9.74	2.59	15.2	5.85		0	1	0		
OL1.12	13.44	-20.14	7.82	0.71	16.44	9.25	2.65	14.54	5.76		0	1	0		
OL1.13	13.64	-20.35	9.17	0.85	15.44	9.72	3.75	13.91	6.54		0	1	0		
OL1.14	13.86	-19.62	8.68	0.81	15.86	9.59	3.1	14.1	6.56		0	1	0		
OL1.15	13.63	-19.85	9.65	0.81	14.73	9.79	3.86	13.84	6.62		0	1	0		
OL1.16	13.78	-19.82	9.08	0.84	15.26	9.71	3.47	14.11	6.8		0	1	0		
OL1.17	13.73	-19.88	9.07	0.82	16.44	10.21	2.79	14.28	6.5		0	1	0		

Abb. 2: Ausschnitt aus dem Microsoft Excel-Datenblatt mit den Datensätzen zur Eingabe in das neuronale Netzwerk für die Klassifikation von Lachsproben nach Produktionsweise. Der Testdatensatz ist blau markiert (s. Text).

Tab. 1: Beispiele zum Vergleich der Ergebnisse der neuronalen Netzanalyse für die Zuordnung der Produktionsweisen von Lachs bei Verwendung verschiedener Eingabedaten

Eingabewerte (kompositionelle Parameter) ^a	Topologie ^b	PE ^c	RE ^c
Wasser + Fett + Asche + Rohprotein + TMA + TVB + SI + FS + Enose + ARE	54~4~3	0	0
Wasser + Fett + Asche + Rohprotein + TMA + TVB	6~4~3	3	3
Wasser + Fett + Asche + Rohprotein + TMA + TVB + SI	8~4~3	0	0
Wasser + Fett + Asche + Rohprotein + TMA + TVB + SI [$\delta^{15}\text{N}$]	7~4~3	0	0
Wasser + Fett + Asche + Rohprotein + TMA + TVB + SI [$\delta^{13}\text{C}$]	7~4~3	2	5
Wasser + Fett + Asche + Rohprotein + TMA + SI	7~4~3	1	0
Wasser + Fett + Asche + Rohprotein + TVB + SI	7~4~3	0	0
Fett + Asche + Rohprotein + TMA + TVB + SI	5~4~3	0	0
Asche + Rohprotein + TMA + TVB + SI	4~4~3	0	0
Rohprotein + TMA + TVB + SI	3~4~3	1	4
Asche + TMA + TVB + SI	3~4~3	1	3
Enose + SI	25~4~3	0	0
Enose	23~4~3	1	0
FS+ SI ^d	15~4~3	0	0
FS[Linolensäure+Eicosensäure+Arachidonsäure] + SI	5~4~3	0	0
FS[Eicosensäure+Arachidonsäure] + SI	4~4~3	0	1
FS[Eicosensäure] + SI	3~4~3	0	1
FS[Arachidonsäure] + SI	3~4~3	3	2

^a Abkürzungen: TMA = Trimethylamin, TVB = flüchtige basische Amine, FS = Fettsäuren, Enos = Electronic Nose, ARE = anorganische Rückstände+Elemente, SI = Stabilisotopen ($\delta^{15}\text{N} + \delta^{13}\text{C}$).

^b Eingabeschicht (input) ~ verdeckte Schicht (hidden layer) ~ Ausgabeschicht (output); vgl. **Abb. 3**

^c Anzahl falscher Zuordnungen bezüglich Testdatensatz (PE = prediction error on test set) bzw. gesamter Trainingsdatensatz (RE = recall error)

^d 4 bis 11 Fettsäuren plus Stabilisotopen ergeben fehlerfreie Klassifizierung (PE/RE = 0)

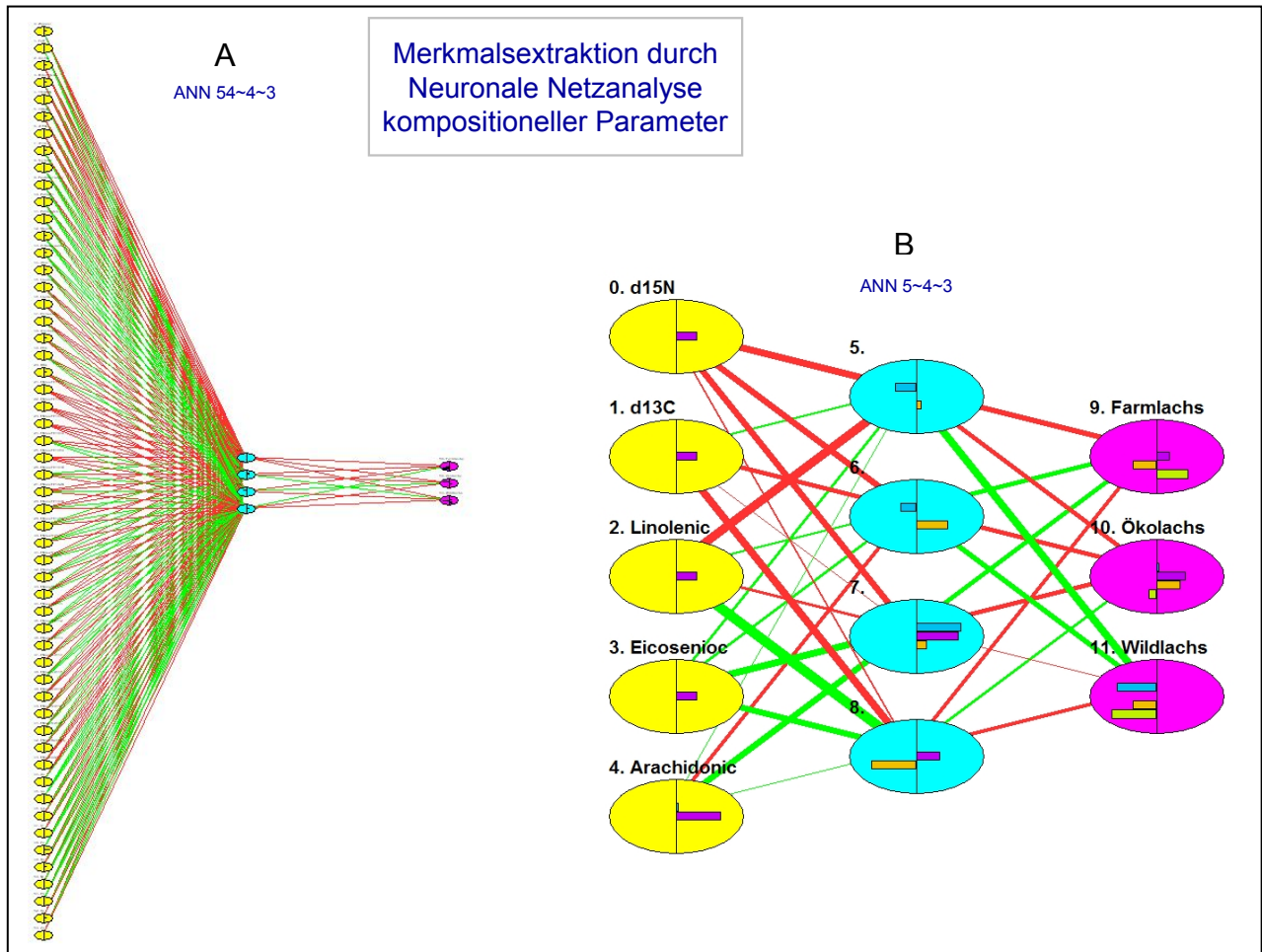


Abb.3: Topologie (Eingabeschicht ~ verdeckte Schicht ~ Ausgabeschicht; Darstellung mit EasyNN-plus v.4.0g) neuronaler Netzwerke für die Klassifikation von Lachsproben nach Produktionsweise. A: Eingabeschicht enthält alle kompositionellen Parameter (s. 1. Zeile in Tab. 1). B: Die Merkmalsextraktion durch neuronale Netzanalyse ermöglicht die korrekte Klassifikation mit wenigen Analysedaten (s. 15. Zeile in Tab. 1).

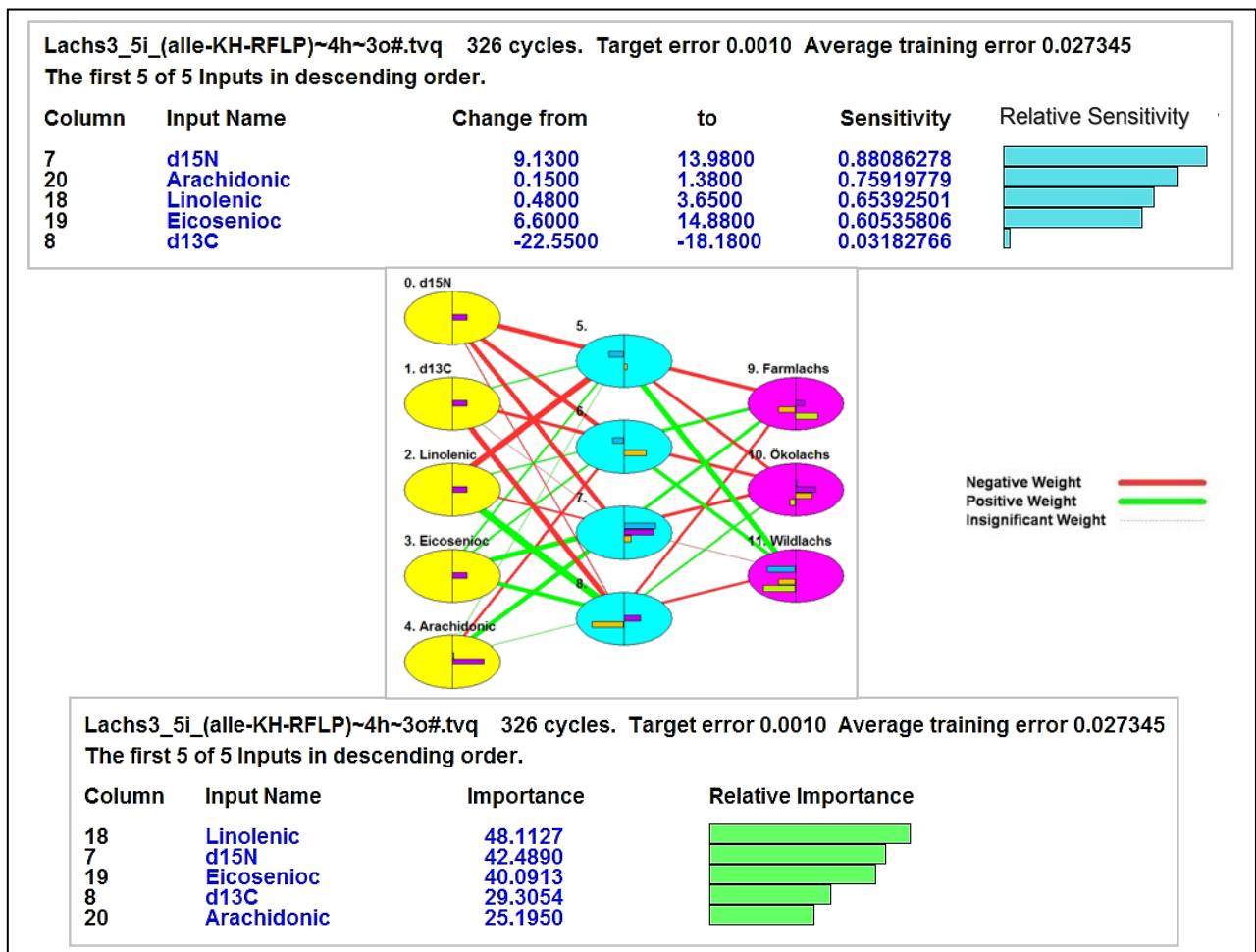


Abb. 4: Analyse (EasyNN-plus v.4.0g) der Einflussfaktoren auf die korrekte Klassifikation von Lachsproben nach Produktionsweise durch das neuronale Netzwerk ANN 5~4~3.
 "Relative Sensitivity": Messgröße für die Ursache-Wirkungs-Beziehung zwischen Input und Output, wobei das Ausmaß der Änderung der Ausgabewertes als Folge der Änderung der Eingabewerte dargestellt wird. "Relative Importance": Messgröße für den Einfluss eines Input-Wertes auf die folgende Neuronen-Schicht im Netzwerk.