

رده بندی سرطان سینه با مقایسه الگوریتم های یادگیری انتشاری پیشین

¹ دانشجوی کارشناسی ارشد مهندسی نرم افزار کامپیوتر، دانشگاه آزاد اسلامی واحد علوم تحقیقات خراسان رضوی،
mahdeian_s@yahoo.com

² دانشجوی کارشناسی ارشد مهندسی نرم افزار کامپیوتر دانشگاه آزاد اسلامی واحد علوم تحقیقات خراسان رضوی،
asmaho202@yahoo.com

چکیده - چندین سال است که سرطان سینه از طریق تکنیک های سیستم یادگیری متفاوت تشخیص داده شده اند. در اینجا ما این رده بندی را با استفاده از شبکه های عصبی مصنوعی Feed Forward ارائه خواهیم داد. الگوریتم های انتشاری پیشین بمنظور یادگیری این شبکه استفاده شده اند. کارایی شبکه با استفاده از مجموعه داده ای سرطان سینه ویسکونسنین برای الگوریتم های یادگیری متفاوت ارزیابی شده اند. موقعه ای که از الگوریتم levenberg marquardt استفاده می شود، بالاترین میزان دقت یعنی 99.28% را خواهیم داشت.

کلید واژه - حداکثر سرطان سینه، الگوریتم انتشاری پیشین، Quasi-Newton.

1- مقدمه

نتیجه اطلاعات را در خود ذخیره کرده و قابل دسترس در مواقع نیاز می سازند. ANN های تکامل یافته در سال های اخیر بمنظور طرح خودکارسازی ANN و اجرای یک جستجوی جهانی بطور گسترده انتشار یافته اند. شبکه های عصبی بمنظور تشخیص سرطان سینه بکار می روند. شبکه های عصبی Feed Forward (FFNN) بطور عمده بمنظور رده بندی مورد استفاده واقع می شوند. این شبکه ها از طریق الگوریتم انتشار پیشین استاندارد بدست آمده اند. این شبکه را شبکه های نظارت شده می گویند که واکنش های مطلوب در مواقع نیاز به آن ها آموزش داده می شود. این الگوریتم ها کارایی رده های آماری بهینه را در شرایط مشکل و نامطلوب تخمین می زنند.

سرطان سینه Wisconsin (WBCD) از طریق محققان گوناگون بمنظور تشخیص پزشکی سرطان سینه در شبکه های عصبی مورد بحث و بررسی قرار گرفته اند. در [2] سرطان سینه با استفاده از شبکه های عصبی Feed Forward از طریق مقایسه نورو ن های مخفی شناسایی شده است. مقایسه کارایی شبکه های ادراکی چند لایه ای از الگوریتم های انتشاری پیشین برای تشخیص سرطان سینه در [5] تحلیل شده است. الگوریتم های یادگیری بکار رفته شامل یادگیری سازگار و متحرک و انتشار پیشین واکنش گر، Quasi-Newton، Levenberg-Marquardt می شود. کارایی های این چهار الگوریتم با الگوریتم انتشار

سرطان سینه بصورت سرطان در بافت سینه رخ می دهد. سرطان سینه یک تومور بدخیم است که از سلول های سینه شروع به رشد می کند. تاکنون این سرطان بیشترین عامل مرگ و میر در بین زنان در کشورهای پیشرفته بده است. مهمترین موثرترین شیوه برای جلوگیری از این خطر تشخیص سریع و بموقع آن است. هرچند تشخیص سریع تر آن به قابلیت تشخیص این نوع از سرطان در مراحل اولیه نیازمند است. تشخیص های اولیه به یک رویه قابل اطمینان و دقیق نیاز دارد تا به پزشکان اجازه تشخیص نوع خوش خیم یا بدخیم بودن آن را بدهند. شناسایی خودکار سرطان سینه مشکلی اساسی و مهم است. در نتیجه پیدا کردن یک روش موثر و دقیق در این نوع از سرطان بسیار حائز اهمیت است. سرطان سینه مسئله ای را که توسط بعضی از دانشمندان در رشته های آماری، داده کاوی و هوش محاسباتی بدست آمده ان، شناسایی می کند.

در حال حاضر شبکه های عصبی مصنوعی ANN بعنوان یک روش موثر برای نمونه برداری، داده کاوی و سیستم یادگیری محسوب می شود. این شبکه های عصبی بطور گسترده بصورت سیستم های پردازش موازی وجود دارن که از عناصر محاسباتی عصبی میانی تشکیل شده اند که توانایی یادگیری دارن و در

پیشین شیبدار استاندارد مقایسه می شوند. شبکه MLP از الگوریتم Levenberg-Marquardt که بالاترین کارایی را دارد بهره می گیرد.

هفتمین ویژگی WBCD بیرناکلی نامیده میشود که شامل 16 متغیر مبهم است. در 2 این شانزده ویژگی موقع ای که WBCD را برای تشخیص سرطان سینه به کار می بریم وجود دارد. شبکه عصبی Feed Forward ساخته شده برای تشخیص سرطان سینه بدون جایگزینی متغیرهای مبهم ارزش یابی شده است. حذف این چند نمونه موثر خواهد بود. در اینجا نتیجه یک رده بندی مسقیم از مجموعه بعد از جایگزینی متغیرهای گم شده که از روش میانه برای مجموعه داده ای wbcd استفاده می کنند ارائه می شود که با چندین نمونه الگوریتم یاد گیری انتشاری پیشین به کار می رود

الگوریتم های یاد گیری با توجه به میزان دقتشان مقایسه می شوند. این مقاله شامل بخش های زیر تقسیم میشود :
بخش دوم : متدولوژی محتوایی بخش
بخش سوم : کاربرد ها و در نهایت نتایج در بخش چهارم ارائه می شود.

2- متدولوژی:

پیشرفت شبکه های عصبی مصنوعی به طور اساسی از طریق تحقیق بر روی سیستم های عصبی بیولوژیکی که شامل شبکه های متراکم نوروها می شود به دست آمده اند در سیستم های عصبی انسانی بیش از 100 نوع نورو وجود دارد ساخت تار اساسی شبکه عصبی در اینجا یک شبکه عصبی چند لایه ای Feed Forward می باشد
تعداد واحد های ورودی متعلق به این خصوصیات نمونه های داده شده می باشد تعداد واحد های خروجی از طریق نتیجه تعیین شده اند. تعداد واحد های پنهان وابسته به مسائل می باشند.

تابع فعال هر واحد یک تابع پیچیده می باشد. شبکه مصنوعی آموزش داده شده از الگوریتم انتشاری پیشین استاندارد استفاده می کند.

شبکه های Feed Forward خروجی هایشان را از طریق نوروهای اختصاصی تغذیه می کنند که به سمت یک یا تعداد بیشتری از نوروها یا لایه ها در شبکه هدایت می شوند. خروجی یک گره از طریق وزن ارتباطی اندازه گیری می شود و

به عنوان ورودی برای گره های بعدی شبکه هدایت می شود. لایه ای ورودی هیچ نقش محاسباتی را بازی نمی کند اما صرفا بردار ورودی را به شبکه انتقال می دهد. لایه ای ورودی و لایه ای پنهان وابسته به وزن ها می باشند بعلاوه لایه پنهان و لایه خروجی هم به وزن های ارتباطی متعلق هستند.

شبکه قابلیت یادگیری از طریق آموزش را دارد این یاد گیری نیازمند مجموعه های بردار های خروجی مرتب و ورودی است. در حین آموزش شبکه مکرر با مجموعه های آموزشی ارائه می شود و وزن ها و مقادیر آستانه در شبکه لحظه به لحظه تا رسیدن به نمونه های ورودی و خروجی مطلوب تنظیم می شود نرون مصنوعی به صورت زیر اجرا می شود.

- 1: سیگنال را از نرون های دیگر دریافت می کند
- 2: هر سیگنال را با توجه به نیروی ارتباطی که همان وزنه است تکثیر می کند
- 3: سیگنال های وزن دار را با هم جمع میکند و به سمتیک تابع فعال سازی هدایت می کند
- 4: خروجی را به دیگر نرون ها میدهد.

معمولا نوروهای لایه ای نهایی می توانند توابع فعال سازی خطی را تا زمانی که نوروهای لایه میانی توابع غیر خطی را پیاده سازی میکنند داشته باشند. به دلیل اینکه اکثریت مشکلات جهان واقعی تفکیک غیر خطی هستند توابع غیر خطی در لایه های میانی برای مسائل پیچیده نمونه سازی شده ضروری هستند.

A: نرمال سازی داده ای

یکی از مهمترین ابزار های شناخته شده که توسط طراحان سیستم های شناسایی خودکار سازی شده به کار می رود به دست آوردن نتایج بهتر است که برای نرمال سازی داده ها مورد استفاده قرار می گیرد در صورتی که یک طراح سیستم محدوده ای یکسان از متغیر ها را برای هر ویژگی ورودی به منظور به حداقل رساندن وقت شبکه در شبکه عصبی از یک ویژگی به ویژگی دیگر می شود

نرمال سازی داده ای همچنین می تواند زمان آموزش را از طریق شروع فرایند آموزشی برای هر خصوصیت در یک مقیاس افزایش دهد.

مجموعه ای ورودی از طریق نرمال سازی مین مکس نرمال شده است که در معادله یک مشاهده می نمایید. که محدوده ای آن بین صفر و یک است.

$$\bar{x} = \frac{(x - x_{min})}{(x_{max} - x_{min})}$$

این نمونه یکی از نمونه های الگوریتم آموزشی شیب دار دسته ای است وزن ها و وقت ها در مسیر شیب منفی تابع اجرایی به روز می شوند.

که X هست مقدار ورودی استاندارد در فرمول x مقدار مشاهده شده است.

x مکرزیموم و x_{min} مینیموم مقادیر مشاهده شده واقعی هستند.

قانون پیش رده بندی زیر در این کار تبیین داده شده است که شامل سه مجموعه ای ضخامت $chnp$ و بیناک لی و میتوسز به صورت زیر است .

اگر ضخامت $chnp$ کوچکتر از هفت و همسانی اندازه سلول کوچکتر از هشت و همسانی شکل سلول کوچکتر از سه و Nucleoli نرمال کوچکتر از نه باشد سپس خوشخیم در غیر این صورت بد خیم .

E: نمونه شیب دار دسته ای متحرک

در به وجود آمدن تقارن سریع تر این سطح شیب دار متحرک به شبکه اجازه میدهد تا هم در سطح شیب دار محلی و هم در آخرین روند سطح خطا عمل کند . عمل کردن به صورت یک فیلتر پایین گذر و متحرک به شبکه اجازه می دهد تا خصوصیات جزی در سطح خطا را نادیده بگیرند . یک شبکه بدون حرکت میتواند در یک مقدار حداقل محلی و سطحی گیر کند .

F: پیوند الگوریتم های شیب دار

الگوریتم انتشاری پیشین اصلی وزن ها را در مسیر شیب دار تنظیم می کند (خنثی سازی سطح شیب دار) مسیر موجود در تابع اجرایی بسار سریع کاهش می یابد این موضوع نشان میدهد که اگر چنین تابع در طول خنثی سازی سطح شیب دار به سرعت کاهش می یابد اما ضرورتا سریعترین همگرایی را تولید نمی کند در الگوریتم های شیب دار پیوندی مطالعه ای در راستای مسیر های پیوندی انجام شده است که به طور کلی همگرایی سریع تری را نسبت به مسیر های شیب دار تولید می کند.

B : جایگزینی متغیر های مبّه یا گم شده که از روش میانی استفاده می کند.

1: پیدا کردن میانه برای بیرناکلی (این ویژگی شامل متغیر گم شده یا مبهم است حد میانه با استفاده از فرمول استفاده می شود)

$$Median = \text{sizeof} \frac{(N + 1)}{2} \text{th item}$$

2: کل مقدار مبهم یا گم شده این ویژگی با این مقدار میانه جایجا شده

به طور کلی چهار مرحله در فرایند آموزشی وجود دارد : 1 جمع آوری داده های آموزشی 2 - ایجاد شبکه شی ای 3-آموزش شبکه 4- شبیه سازی واکنش شبکه به ورودی های جدید چندین متغیر برای الگوریتم انتشاری پیشین وجود دارند که در اینجا توضیح داده شده است.

C: آموزش دسته ای

در این روش وزن ها و وقت شبکه تنها بعد از اتمام مجموعه آموزشی به کار رفته در شبکه به روز می شوند شیب ها در هر نمونه آموزشی محاسبه شده تا تغییرات موجود در وزن ها و وقتهای شبکه را تامین کنند.

G: الگوریتم های Quasi-Newton

روش نوتون جایگزینی برای روش های شیب دار پیوندی به منظور بهینه سازی سریع تر است مجموعه ای از الگوریتم های مبنی بر روش نوتون وجود دارد که هیچ کدام از آنها به محاسبه ای مشتقات نیاز ندارند این روش Quasi-Newton یا secant نامیده می شوند آنها یک ماتریس هسین تقریبی را در هر تکرار الگوریتم به روز می کند . این بروز رسانی به عنوان یک تابه شیب دار محاسبه شده است .

H: Levenberg-Marquardt

این الگوریتم شبیه روش قبلی به منظور دست یافتن به سرعت آموزشی سطح دوم بدون در نظر گرفتن ماتریس هسین طراحی شده است.

D: نمونه شیب دار دسته ای

I: انتشار resilient back

شبکه های چند لایه ای به طور عمده از توابع انتقالی حلقوی - پیچیده در لایه های پنهانی استفاده میکنیم این توابع اغلب توابع اسپاشینگ نامیده می شود به دلیل اینکه آنها یک محدوده ورودی نامحدود را درون یک ورودی محدود فشرده سازی میکنند .

این توابع در واقع سطوح شیب دار را با زیاد شدن میزان ورودی به صفر می رسانند موقعه ای که شما از تابع حلقوی و سطح شیب دار برای آموزش شبکه چندلایه ای بهره می گیرید دچار مشکل می شوید بدلیل اینکه شیب می تواند بسیار کوچک باشد و در نتیجه موجب تغییرات بسیار محدود در وزن و زمان شبکه شود. با این وجود وزن ها و وقت ها بسیار دورتر از مقادیر بهینه هستند. هدف الگوریتم آموزشی انتشاری پیشین واکنش گرا، از بین بردن این تاثیرات مضر در اندازه های مشتقات جزئی است. تنها علامت مشتقات جزئی بمنظور تعیین شیوه به روز رسانی وزن بکار می رود و اندازه آن هیچ تاثیری در به روز رسانی وزن نداشته است. تغییر بوجود آمده در اندازه وزن با توجه به یک مقدار به روزرسانی مجزا مشخص شده است.

الگوریتم انتشاری پیشین و روش آموزشی نظارت شده در این پروژه بکار رفته اند. هدف از آموزش تنظیم وزن هاست تا موقع ای که خطا بین خروجی مطلوب و خروجی واقعی کاهش یافته اندازه گیری شده است. آموزش موقع ای که به یک مقدار نسبتاً موثر رسید متوقف می شود به منظور تحلیل جعبه ابزار شبکه عصبی داده ای که در نرم افزار مطلب قابل دسترس است به کار می رود الگوریتم مورد نظر در این تحقیق به صورت زیر است.

1: مجموعه ای داده ای اجرا می شود و مقادیر مبهم یا گم شده با استفاده از روش میانه جایگزین میشود

2: داده ها با استفاده از نرمال سازی مین مکس نرمال می شوند

3: با استفاده از الگوریتم انتشاری پیشین یک شبکه ساخته می شود که با یک لایه ای ورودی و یک لایه ای پنهان و یک

لایه ای خروجی است ،در لایه ای ورودی 9نرون در لایه ای پنهان 6 نرون و 1 نرون در لایه ای خروجی است
4: یک مقدار تصادفی برای وزن و وقت ورودی تعیین می شود

5: 80%به منظور آموزش و 20%از آن برای تست در نظر گرفته می شود

6:با استفاده از الگوریتم انتشاری پیشین و با توجه به الگوریتم های آموزشی شبکه آموزش و تست می شود

7: نتایج الگوریتم های آموزشی با توجه به دقت آنها مقایسه می شوند

در ابتدا شما باید یک الگوی درست برای اندازه ای صفحه تعیین کنید این الگو برای خروجی در اندازه ای صفحه us-letter تعیین شده است اگر شما از صفحه A4استفاده میکنید لطفا این فایل را ببینید و فایل msw A4 را دانلود کنید.

3- یک کاربرد

3-1: پایگاه داده ای سرطان سینه ای wisconsin

این پایگاه داده در اصل توسط دکتر ویلیام والبگ به دست آمده و توسط تعدادی از محققان در نمونه برداری و ماشین یادگیری مورد استفاده قرار گرفته است پایگاه داده ای موجود در uci شامل 699 نمونه است مجموعه ای داده ای اصلی شامل 11 ویژگی می شود که هر دو نمونه ای شماره آی دی و برچسب کلاس را در بر می گیرد این دو نمونه در مجموعه ای داده ای اصلی که در این کاربرد به کار رفته است حذف شده است کلاس هر نمونهدر بر گیرنده ای نمونه های خوش خیم و بد خیم است 9 ویژگی باقی مانده 9 خصوصیت سیتولوژی fina سینه را همانطور که در جدول یک نشان داده شده است را ارائه می دهد . این خصوصیات از یک تا ده از خوش خیم به بد خیم ارزش گذاری شده اند .

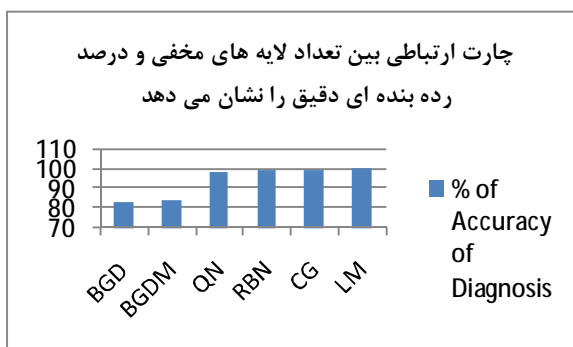
NO	Attribute	Domain
1	Clump thickness	1-10
2	Uniformity of cell size	1-10
3	Uniformity of cell shape	1-10
4	Marginal Adhesion	1-10
5	Single Epithelial cell size	1-10
6	Bare Nuclei	1-10
7	Bland Chromatin	1-10
8	Normal Nucleoli	1-10
9	Mitoses	1-10
10	Class	2 for benign and 4 for malignant

مجموعه ای داده ای به دو مجموعه تقسیم شده است :
 الگوریتم آموزشی لونبرگ بهترین نتیجه ای تشخیص را به دست آورده است .
 مجموعه آموزشی و تست

جدول ۱: عملکرد الگوریتم های آموزش.

Name of the Training Algorithm	% of Accuracy of Diagnosis
Batch Gradient Descent (BGD)	83.27
Batch Gradient Descent with Momentum (BGDM)	84.39
Quasi Newton (QN)	98.42
Resilient Back Propagation (RBP)	98.63
Conjugate Gradient (CG)	98.99
Levenberg Marquardt (LM)	99.28

شکل یک چارت ارتباطی بین تعداد لایه های مخفی و درصد رده بنده ای دقیق را نشان می دهد.



مجموعه تست از تریق هیچ شبکه عصبی در طول مرحله آموزشی دیده نشده اند که تنها به منظور تست ترمیم گروه های شبکه عصبی بعد از آموزش به کار می روند ما 80% از نمونه را برای مجموعه آموزشی و 20% از باقی مانده آن را برای مجموعه تست به کار برده ایم . 3 لایه در شبکه انتشاری پیشین که شامل یک لایه ای ورودی حاوی واحد های 9 واحد و یک لایه ای واحد مخفی حاوی 6 واحد و یک لایه ای خروجی شامل یک واحد می شود وجود دارند ارزش واحد در لایه ای خروجی لایه ای ورودی را به صورت نرمال یا غیر نرمال نشان داده اند در ابتدای آموزش تمام وزن های ارتباطی در شبکه با مقادیر تصادفی ارزش گذاری شده اند . کل بردار های ورودی نرمال سازی شده اند بنا بر این به ترتیب مقدار حداقل و حداکثر آن صفر و یک می باشند . تمام محاسبات با استفاده از مطلب ورژن هفت پیاده سازی شده اند تابع newff به منظور ایجاد شبکه انتشاری Feed Forward back به کار رفته است

طراحی در این کاربرد ها شامل واحد های مخفی قهوه ای و یک واحد خروجی purlin می شود . نسبت یاد گیری 0.7 استفاده شده است تعداد حداکثر دوره های مجاز 1000 بوده است

جدول 2 تاثیر الگوریتم های یاد گیری بر میزان تشخیص الگوریتم انتشاری پیشین نشان می دهد از این جدول در میابیم که شبکه عصبی Feed Forward از الگوریتم انتشاری پیشین با

[9] Paulin F. and Santhakumaran A., "Extracting Rules from Feed Forward Neural Networks for Diagnosing Breast Cancer" CiiT

International Journal of Artificial Intelligent Systems and Machine Learning, vol. 1, No. 4, July 2009, pp. 143-146

[10] Punitha A., Sumathi C.P. and Santhanam T., "A Combination of Genetic Algorithm and ART Neural Network for Breast Cancer

Diagnosis" Asian Journal of Information Technology 6 (1):112-117, 2007, Medwell Journals, 2007.

[11] Renato De Leone, Rosario Capparuccia and Emanuela Marelli, "A Successive Overrelaxation Backpropagation Algorithm for Neural-

Network Training" IEEE Transactions on Neural Networks, vol. 9, No. 3, May 1998, pp. 381-388

F.Paulin et al. / International Journal on Computer Science and Engineering (IJCSE)

ISSN : 0975-3397 Vol. 3 No. 1 Jan 2011 331

[12] Rudy Setiono and Huan Liu, "Neural-Network Feature Selector" IEEE Transactions On Neural Networks, vol. 8, No. 3, May 1997, pg

664-662

[13] Tuba Kiyani And Tulay Yildirim, "Breast Cancer Diagnosis Using Statistical Neural Networks" Istanbul University, Journal Of

Electrical And Electronics Engineering, Year 2004, vol. 4, Number 2, pp.1149-1153

[14] Sudhir D. Swarkar, Ashok Ghatol, Amol P. Pande, "Neural Network Aided Breast Cancer Detection and Diagnosis Using Support

Vector Machine" Proceedings of the International conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006, pp. 158-163.

[15] UCI Machine Learning Repository. [<http://archive.ics.uci.edu/ml/>]. Irvine, CA: University of California, Center for Machine Learning and Intelligent Systems.

[16] Włodzisław Duch and Rafał Adamczak and Krzysztof Grabczewski, "A New methodology of Extraction, Optimization and

Application of Crisp and Fuzzy Logic Rules" IEEE Transactions On Neural Networks, vol. 12, No. 2, March 2001, pp. 227-306

4- نتیجه‌گیری

در این تحقیق یک شبکه عصبی Feed Forward ساخته شد و الگوریتم انتشاری back برای آموزش شبکه به کار برده شد. الگوریتم مورد نظر روی مسائل واقعی (مسئله تشخیص سرطان سینه) wiscosin مورد آزمایش قرار گرفت در اینجا 6 الگوریتم یادگیری به کار برده شد. بهترین آنها لونبرگ بود با میزان 99.28 % شناخته شد در این تشخیص نرمال سازی مینمکس به کار برده شد.

مراجع

[1] Abraham, A. "Meta learning evolutionary artificial neural networks", Neuro computing, 2004, 56, pp.1-38.

[2] Anupam Shukla, Ritu Tiwari and Prabhdeep Kaur, "Knowledge Based Approach for Diagnosis of Breast Cancer" IEEE International

Advance Computing Conference, Patiala, India, March 2009, pp. 6-12

[3] Basheer I.A., and Hajmeer M. "Artificial neural networks: fundamentals, computing, design, and application" J. Microbiol. Methods 43 (2000) 3-31.

[4] Bothorel S., Meunier B.B., and Muller S.A., "Fuzzy logic based approach for semi logical analysis of micro calcification in

mammographic images" International Journal of Intelligent System, 12 (1112): 819-848, 1997.

[5] Esugasini S, Mohd Yusoff Mashor, Nor Ashidi Mat Isa and Nor Hayati Othman (2005), Performance Comparison for MLP Networks

Using Various Back Propagation Algorithms for Breast Cancer Diagnosis, Knowledge-Based Intelligent Information and Engineering

Systems, Lecture Notes in Computer Science, 3682, pp. 123-130.

[6] Furundzic D., Djordjevic, and Bekic A. J., "Neural Networks approach to early breast cancer detection", Systems Architecture, 44:

617- 633, 1998.

[7] Jun Zhang MS, Haobo Ma Md MS, "An Implementation of Guildford Cytological Grading System to diagnose Breast Cancer Using

Naïve Bayesian Classifier", MEDINFO 2004, M.Fieschi et al. (Eds),Amsterdam:IOS Press

[8] Kamruzzaman S.M. and Monirul Islam Md, "Extraction of Symbolic Rules from Artificial Neural Networks" Proceedings of world

Academy of science, Engineering and Technology, vol. 10, Dec. 2005, ISSN 1307-6884