



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

## ارزیابی کارایی روش های درخت تصمیم برای رده بندی پوشش زمین

### چکیده

انتخاب یک الگوریتم طبقه بندی معمولاً مبتنی بر چندین عامل است، که از میان آن ها دسترسی پذیری نرم افزار، سهولت استفاده، و عملکرد در اینجا بواسطه دقت طبقه بندی کلی ارزیابی شده اند. رویه ماکسیمم مشابهت (ML) برای بسیاری از کاربران الگوریتم انتخاب است چون آماده استفاده است و این حقیقت که مستلزم فرایند آموزش بسط یافته نمی باشد. شبکه های عصبی مصنوعی (ANN ها) اکنون به طور گسترده توسط محققان استفاده می شوند، اما از کاربردهای عملیاتی آن ها بواسطه نیاز به کاربر برای تخصیص پیکربندی معماری شبکه برای ارائه مقادیری برای تعدادی از پارامترها جلوگیری شده است، که هر دو روی عملکرد تأثیر می گذارند. ANN نیز مستلزم فاز آموزش بسط یافته می باشد.

در چند سال اخیر، استفاده از درخت های تصمیم (DTها) برای رده بندی داده هایی که از راه دور حس شده اند افزایش یافته اند. طرفداران این روش ادعا می کنند که این درخت مزایای متعددی نسبت به الگوریتم های ML و ANN دارند. DT از نظر محاسباتی سریع است، و هیچ فرضیه آماری ایجاد نمی کند، و می تواند داده هایی را کنترل کند که در مقیاس های سنجش مختلف نشان داده شده است. نرم افزار لازم برای پیاده سازی DT ها به راحتی از اینترنت قابل دسترسی است. هرس کردن DT ها می تواند باعث کوچک تر شدن آن ها و تفسیر ساده تر و راحت تر آن ها بشود، در حالی که استفاده از تکنیک های تقویت می تواند باعث بهبود عملکرد شود.

در این مطالعه، مجموعه تست جداگانه و مجموعه های داده ای آموزشی از دو حوزه جغرافیایی مختلف و دو سنسور مختلف برای ارزیابی عملکرد DT های تک متغیره و چند متغیره برای طبقه بندی پوشش زمین استفاده شده اند. عوامل در نظر گرفته شده به شرح زیر هستند: اثرات تغییرات در اندازه مجموعه داده های آموزشی چند بعدی بود فضای ویژگی، همراه با تأثیر توییت، مقیاس های انتخاب ویژگی، و هرس کردن. سطح دقت رده بندی DT تک متغیره با نتایج ANN پس انتشار و رده بندی کننده های ML مقایسه شده است. نتایج ما نشان می دهد که

عملکرد DT تک متغیره در مقایسه با عملکرد دیگر رده بندی کننده ها قابل قبول است، به جز با داده های دارای ابعاد بالا. دقت رده بندی به طور خطی با اندازه مجموعه داده های آموزشی تا حد 300 پیکسل در هر رده در این مورد افزایش می یابد. DT های چند متغیره عملکرد بهتری نسبت به DT های تک متغیره ندارند. در حالی که تقویت باعث افزایشی در دقت رده بندی بین 3 و 6 درصد می شود و به نظر می رسد که روش های انتخاب ویژگی با توجه به افزایش های ایجاد شده در دقت تنظیم شود. اما، نه DT تک متغیره و نه DT چند متغیره عملکرد مشابه با ANN یا رده بندی کننده های ML با ابعاد دارای ابعاد بالا دارند.

**کلیدواژگان:** درخت تصمیم، پوشش زمین، الگوریتم طبقه بندی

## 1- مقدمه

در سه دهه قبل شاهد توسعه های مداوم در حوزه تشخیص الگو بودیم. تحقیق در جنبه های الگوریتمی تشخیص الگو در کنار توسعه ابزارهایی پیش رفته است که قادر به تولید مقادیر بالایی از داده ها می باشند، شامل تصاویر با وضوح فضایی و طیفی بهتر. پس از 30 سال حسگری از راه دور ماهواره ای از سطح زمین، کاربرانی که از راه دور داده ها را حس می کنند اکنون به الگوریتم های آماری پیچیده و عصبی / ارتباط گرا برای رده بندی های فازی و سخت داده های آن ها دسترسی دارند.

هر دو روش آماری و عصبی / اتصال گرا محدودیت هایی دارند. روش های آماری تکیه بر این فرضیه دارند که احتمالات عضویت در رده می تواند با یک تابه چگالی احتمال خاص مدلسازی شود. در اغلب موارد، توزیع اقلیدسی انتخاب شده است، چون بواسطه آماره ترتیب اول و ترتیب دوم توصیف شده است، یعنی بردارهای میانگین رده و ماتریس های کوواریانس رده. اگر اندازه مجموعه آموزشی ثابت باشد، دقت تخمین ای عناصر بردار میانگین رده نمونه و ماتریس کوواریانس رده نمونه با افزایش تعداد ویژگی ها (ابعاد) کاهش می یابند. این فرضیه که داده ها در هر رده از مدل نرمال چند متغیره تبعیت می کنند تحلیل وقفه یا داده های مقیاس نسبت را محدود می کند.

به نظر می رسد که روش های عصبی / اتصال گرا به خوبی با مجموعه های داده ای آموزشی کار می کنند که اندازه کوچکتری نسبت به آن هایی دارند که در روبه های آماری لازم هستند. از طرف دیگر، زمان های آموزش شبکه می

توانند طولانی باشند، در حالی که انتخاب طرح معماری شبکه (از نظر تعداد لایه های پنهان و نورون ها در هر لایه) و مقادیر پارامترهای نرخ یادگیری رو به جلو نیستند. بر خلاف روش های آموری، مشی عصبی / اتصال گرا هیچ فرضیه ای در رابطه با توزیع بسامد آماری داده ها با مقیاس های اندازه گیری (ارزیابی) ویژگی هایی که در تحلیل استفاده شده اند ارائه نمی کنند. معمول ترین الگوریتم عصبی / اتصال گرای استفاده شده پرسپترون چند لایه پس انتشار است که در این مطالعه استفاده شده است.

رده بندی کننده های درخت تصمیم بطور گسترده در گروه حسگری از راه دور به گستردگی روش های آماری یا روش های عصبی / اتصال گرا مورد استفاده قرار نگرفته اند. مزایایی که درخت های تصمیم ارائه می کنند شامل توانایی کنترل داده های ارزیابی شده با مقیاس های مختلف، فقدان هر گونه فرضیه ای در رابطه با توزیع بسامدهای داده ها در هر یک از رده ها، انعطاف پذیری، و توانایی کنترل روابط غیر خطی میان ویژگی ها و رده ها می باشد. در مقابل شبکه های عصبی، درخت های تصمیم می توانند به سرعت آموزش داده شوند، و اجرای آن ها سریع است. آن ها می توانند برای انتخاب / کاهش و همچنین به مقاصد رده بندی مورد استفاده قرار گیرند. در نهایت، تحلیل گر می تواند یک درخت تصمیم را تفسیر کند. این جعبه سیاه نیست، مثل شبکه عصبی، یعنی کارهای پنهانی که از دید پنهان شده اند.

دقت رده بندی کلی در این جا برای اندازه گیری عملکرد روش های مختلف استفاده شده اند. سطح دقت رده بندی که در یک مورد خاص بدست آمده است بسته به تعدادی از عوامل می باشد، شامل طبیعت مسأله رده بندی از نظر پیچیدگی محدودیت های تصمیم که رده ها را در فضای ویژگی، اندازه نمونه آموزشی، کفایت داده های آموزشی در توصیف ویژگی های رده های منتخب، چند بعدی بودن داده ها و ویژگی های رده بندی استفاده شده جدا می کنند. ما تمام این مشکلات را در این مقاله در نظر نمی گیریم. اما نتایج تحلیل های ما در داخل خودشان قابل مقایسه اند، چون همان مجموعه های داده ای آموزش و تست برای هر سه رده بندی در دو حوزه مختلف مطالعه استفاده شده اند. از این رو آزمودن عملکرد نسبی رده بندی کننده های مختلف و سازگاری این مقایسات میان مجموعه های داده ای با مشخصه های نامشابه از نظر زمینه حوزه مطالعه و طبیعت سیستم تصویربرداری استفاده شده ممکن می باشد.

مقاله به شرح زیر سازمان دهی شده است. در بهش دو دو مجموعه داده ای تست شرح داده شده اند که در این مطالعه استفاده شده اند. جزئیات خلاصه سه رده بندی کننده در بخش 2 ارائه شده اند. اثرات اندازه مجموعه آموزشی، چند بعدی بودن داده ها، روش های انتخاب ویژگی، هرس کردن و تقویت عملکرد رده بندی کننده های DT در بخش 4 در نظر گرفته شده اند. تحلیل تطبیقی کوتاه عملکرد نسبی DT، شبکه های عصبی مصنوعی (ANN ها) و ماکسیمم مشابه رده بندی کننده ها در بخش 5 داده شده است که به دنبال آن خلاصه ای از نتایج ارائه شده است.

## 2- مجموعه های داده ای تست

دو مجموعه داده ای مغایر استفاده شده اند. اولین مجموعه تصویری با رزولوشن متوسط از بخشی از شرق انگلستان و نزدیک شهر Littleport است. این حوزه نسبتاً هموار است و عمدتاً به کشاورزی و زراعت شدید اختصاص یافته است. تصویر در 19 ژوئن 2000 جمع آوری شده است. هفت نوع پوشش زمین اصلی شناسایی شده اند، یعنی، گندم، سیب زمینی، چغندر قند، پیاز، نخود فرنگی، کاهو و باقلا. خروجی چاپ شده داده های زمینه رسمی برای سال 2000 از کشاورزان و عوامل نماینده آن ها جمع آوری شده اند و دیگر بخش های ناحیه روی زمین برای همگذاری داده های مرجع بررسی شده اند. تصویر جزئی متشکل از 307 پیکسل در 330 پیکسل پوشش دهنده حوزه سود از تصویر EMT+ برای تحلیل های متعاقب استخراج شدند. باند 6 ETM+ حذف شدند.

داده های ابر طیفی بدست آمده با استفاده از طیف سنج تصویر DAIS 7915 در 29 ژوئن 2000 از مجموعه داده های تست ثانویه به دست آمدند. رزولوشن فضایی داده های DAIS 5 متر است و مقیاس ها در 72 نوار طیفی در نواحی عینی و موج کوتاه طیف ایجاد شده اند. حوزه مطالعه در ناحیه La Mancha Alta قرار گرفته است که حوزه تقریباً 8000 کیلومتر مربع را پوشش می دهد و در جنوب مادرید اسپانیا قرار گرفته است. این ناحیه دارای زمین مرطوب نیمه خشک با کشاورزی زمین خشک و آبیاری می باشد. هشت نوع پوشش زمین مختلف شناسایی شده اند. Subimage ای با اندازه 512×512 پیکسل پوشش دهنده حوزه سود استخراج شده است. از 72 نوار

موجود در ناحیه عینی و موج کوتاه موجودند، زیر مجموعه متشکل از 65 نوار انتخاب شدند، و به عنوان بررسی بصری نشان دادند که هفت نوار از اثرات برهنه سازی افقی شدید رنج می برند.

روش های نمونه گیری تصادفی برای جمع آوری مجموعه های داده ای آزمایشی و تست مجزا در هر دو حوزه از داده های جمع آوری شده از مشاهدات میدانی و در مورد حوزه مطالعه Littleport از سوابق کشاورزی رسمی استفاده شدند. پیکسل های جمع آوری شده با نمونه گیری تصادفی به دو زیر مجموعه تقسیم شدند، که یکی از آن ها برای آموزشی و دومی برای آزمایش رده بندی کننده ها استفاده شدند، تا هر پیشقدر منتجه ای از استفاده از همان مجموعه پیکسل ها برای مجموعه های داده ای تست و آزمایشی حذف شود. علاوه بر این چون همان مجموعه های داده ای تست و آموزشی برای هر رده بندی کننده استفاده شده اند، از هر تفاوت منتجه از تغییرات نمونه گیری جلوگیری شده است.

### 3- روش ها

ماکسیمم مشابهت، شبکه عصبی پس انتشار چند لایه، و رویه های درخت تصمیم در این مطالعه استفاده شده اند. خلاصه موجزی از ویژگی های هر یک از این رده بندی کننده ها در این بخش داده شده است.

#### 3-1- رده بندی کننده ماکسیمم مشابهت

همان که پیاده سازی شده است، رویه ML مبتنی بر فرضیه ای است که در آن اعضای هر رده از توزیع فرکانس اقلیدوسی در فضای ویژگی استفاده می کنند. ML یک روش مبتنی بر پیکسل است و می تواند به شرح زیر تعریف شود: یک پیکسل با بردار ویژگی مشاهده شده متناظر  $X$  به رده  $C_j$ ،  $N$  رده اختصاص یافته است، اگر

$$g_j(\mathbf{x}) > g_k(\mathbf{x}) \text{ for all } j \neq k, \text{ with } j, k = 1, \dots, N.$$

برای توزیع گاوسی چند متغیره، تابع افترافی  $g_k(\mathbf{x})$  با فرمول زیر داده شده است:

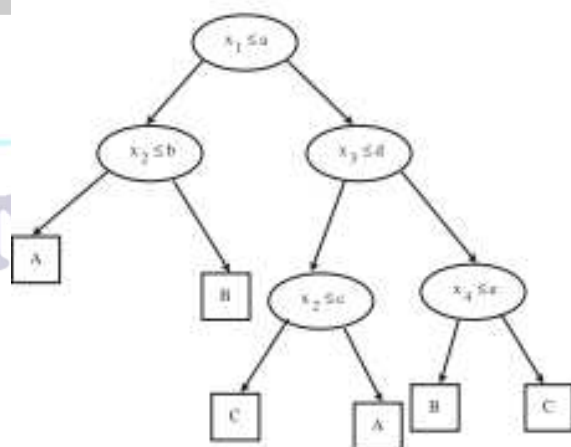
$$g_k(\mathbf{x}) = \ln(p(\mathbf{x} | c_j)) = \ln \hat{\Sigma}_k + (\mathbf{x} - \hat{\mu}_k)^T \hat{\Sigma}_k^{-1} (\mathbf{x} - \hat{\mu}_k)$$

که در آن  $\hat{\mu}_k$  و  $\hat{\Sigma}_k$  بردار میانگین نمونه و ماتریس کوواریانس نمونه برای رده  $k$  هستند. پیاده سازی الگوریتم ML شامل تخمین بردارهای میانگین رده و ماتریس های کوواریانس از داده های آموزشی انتخاب شده از نمونه های معینی از هر رده خاص هستند. تابع  $g_i(X)$  برای ارزیابی عضویت یک پیکسل ناشناخته برای رده  $j$  استفاده شده است. پیکسل به رده ای تخصیص یافته که برای آن بالاترین مقدار احتمال عضویت وجود دارد.

## 2-3- رده بندی کننده شبکه عصبی مصنوعی

پر استفاده ترین مدل شبکه عصبی در کاربردهای حس گری از راه دور پرسپترون چند لایه پس انتشار است. طرح آن از یک لایه ورودی، حداقل یک لایه پنهان، و یک لایه خروجی تشکیل می شود. لایه های پنهان و خروجی از مجموعه هایی از واحدهای پردازش غیر خطی، یا نورون ها ساخته شده اند و ارتباطاتی میان نورون ها در لایه های بعدی دارند که حامل وزن های متناظر می باشند.

شکل 1.



اطلاعات تنها در جهت جلو حمل می شوند، یعنی، از لایه ورودی به اولین لایه پنهان، یا از یک لایه پنهان به لایه پنهان بعدی یا لایه خروجی. پردازش غیر خطی با اعمال تابع فعال سازی در ورودی های جمع بندی شده در هر نورون انجام شده است. شبکه با استفاده از پس انتشار آموزش داده شده است که از الگوریتم Gradient Descent برای به حداقل رساندن خطا میان برچسب شناخته شده ای از پیکسل آموزشی و خروجی برچسب با شبکه برای آن پیکسل استفاده می کند. هر عضو از یک مجموعه پیکسل های آموزشی مکررا در شبکه ارائه شده اند، و خطا از لایه

خروجی به لایه ورودی پس فرستاده شده است. وزن ها روی مسیر بازگشت به عقب از طریق شبکه طبق یک قانون به روز رسانی و نرخ یادگیری به روز رسانی شده اند. ANN ها منحصراً بواسطه مشخصه های واحدهای پردازش آن ها و قانون آموزشی یا یادگیری منتخب تخصیص یافته اند. توپولوژی شبکه، یعنی تعداد لایه های پنهان و تعداد نوروں ها در هر لایه، تدثیر قابل توجهی روی عملکرد دارد. هیچ راهنمای مشخصی برای تعیین این وجود ندارد که معماری شبکه برای انتخاب هر یک از مقادیر ابتدایی پارامترهای تأمین شده توسط کاربر استفاده شده است یا نه که به عنوان مثال عملکرد رویه کمینه سازی خطا را کنترل می کنند. تنظیمات توصیه شده توسط Kavzoglu (2001) در اینجا استفاده شده اند.

### 3-3- رده بندی کننده های درخت تصمیم

بر خلاف رده بندی کننده های آماری و عصبی / اتصال گرای مرسوم که از تمام ویژگی های موجود به طور همزمان استفاده می کنند و تصمیم عضویت منفردی برای هر پیکسل می گیرند، DT از مشی چند مرحله ای یا ترتیب برای مشأله تخصیص برچسب استفاده می کند. فرایند برچسب گذاری به عنوان زنجیره ای از تصمیمات ساده مبتنی بر نتایج تست های ترتیبی در نظر گرفته شده است نه تصمیمی منفرد و پیچیده. مجموعه دنباله های تصمیم شاخه هایی از DT با تست هایی که در گره های استفاده شده اند، تشکیل می دهد. برگ ها نشان دهنده برچسب ها هستند. ساخت DT شامل تقسیم بندی بازگشتی یک مجموعه داده آموزشی است که به زیر مجموعه های یکنواخت بر مبنای تست های بکار رفته در یک یا چند مقدار ویژگی تقسیم شده است. این تست ها با گره ها نشان داده شده اند. DT تک متغیره یک تست را در یک ویژگی منفرد در یک زمان به کار می برد، در حالی که DT چند متغیره از یک یا چند ویژگی به صورت همزمان استفاده می کند. برچسب هایی به گره های پایانی با استفاده از استراتژی تخصیص اختصاص یافته اند. در هر وهله، DT ها به صورت دستی طراحی شده است، که از پلات های طیفی استفاده می کنند. در دهه قبل، روش های اتوماتیک طرح درخت تصمیم توسعه یافته اند. در این مطالعه، دو الگوریتم DT تک متغیره، یعنی C4.5 و See 5.0 ، و یک الگوریتم چند متغیره استفاده شده اند.

### 1-3-3- درخت تصمیم تک متغیره



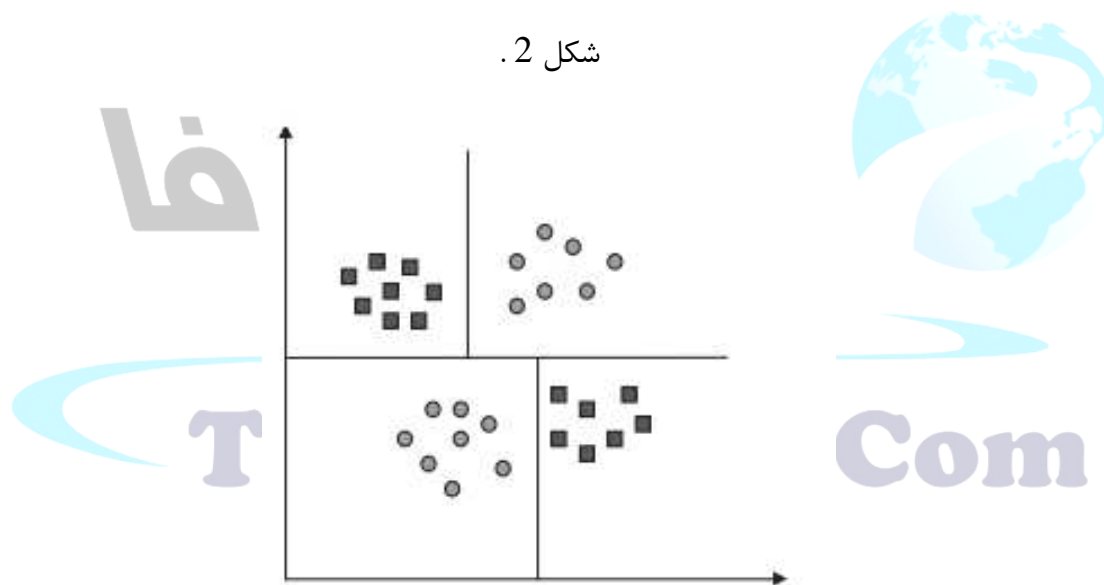
DT تک متغیره درختی است که در آن محدودیت های تصمیم در هر گره درخت با خروجی تست به کار برده شده در یک ویژگی منفرد تعریف شده که در هر گره داخلی ارزیابی شده است. بر مبنای خروجی تست، داده ها به دو یا چند زیرمجموعه تقسیم شده اند. هر تست باید تعداد خروجی های مجزا داشته باشد. رده بندی DT تک متغیره با تقسیم بندی بازگشتی داده های ورودی تا زمانی ادامه می دهد که به یک گره برگ برسیم و برچسب رده متناظر با گره برگ پس از آن به آن مشاهده تخصیص می یابد. مشخصه های محدودیت های تصمیم در یک DT تک متغیره به صورت تجربی از داده های آزمایشی تخمین زده شده اند. در مورد داده های پسوخته، تستی به شکل  $x_i < C$  در هر گره داخلی DT انجام شده است، که در آن  $x_i$  مقیاسی در فضای ویژگی است و  $C$  آستانه تخمین زده شده از توزیع  $x_i$  است. مقدار  $C$  با استفاده از مقیاس هدفی تخمین زده شده که نامشابهت ها را به حداکثر می رساند یا شباهت های گره های فرزند را به حداقل می رساند و در هر لحظه از یک ویژگی استفاده می کند. تعداد روش انتخاب ویژگی در مقالات شرح داده شده اند. روش هایی که بیش تر از همه مورد استفاده قرار می گیرند، سود اطلاعاتی، نرخ سود اطلاعاتی، شاخص Gini، و مقیاس مجذور خی می باشند. چون هر تست در DT تک متغیره مبتنی بر یک ویژگی منفرد است، محدود به محوری است که نشان دهنده ویژگی منتخب می باشد.

### 2-3-3- درخت های تصمیم چند متغیره

اگر مکان های محدودیت های تصمیم در فضای ویژگی را بتوان به درستی و تنها از نظر ترکیب های ویژگی به جای دنباله هایی از ویژگی های منفرد تعریف کرد، DT تک متغیره عملکرد ضعیفی خواهد داشت. در چنین مواردی، مجموعه تقسیمان مجاز می توانند بسط یابند تا تلفیق هایی خطی از ویژگی ها شامل نمایند. مجموعه ای از توابع افتراقی خطی در هر گره داخلی از یک DT چند متغیره به همراه ضرایب برای تابع افتراقی خطی در هر گره داخلی که از داده های آموزشی تخمین زده می شود، تخمین زده شده است. تست تقسیم کردن در هر گره فرم  $\sum_{i=1}^n a_i x_i \leq c$  را دارد که در آن  $x_i$  نشان دهنده برداری از مقیاس ها روی  $n$  ویژگی منتخب است،  $a$  بردار ضرایب یک تابع افتراقی خطی است و  $C$  مقدار آستانه است. Utgoff و Brodley (1992) دریافتند که DT های چند متغیره فشرده تر هستند و قادر به تولید رده بندی های دقیق تری نسبت به DT های تک متغیره می باشند. هر چه

پیچیدگی الگوریتم های DT چند متغیره نسبت به DT تک متغیره بیشتر باشد عواملی معرفی می شوند که روی عملکرد آن ها تأیر می گذارند. ابتدا، هر یک از تعداد الگوریتم های متفاوت می توانند برای تخمین قانون تقسیم بندی در گره های داخلی استفاده شوند و عملکرد نسبی ای نروش های می توانند بسته به طبیعت داده ها و پیچیدگی مسأله رده بندی متفاوت باشند. ثانیاً، چون هر بخش در هر گره داخلی یک DT چند متغیره مبتنی بر یک یا چند ویژگی است، چندین الگوریتم مختلف می توانند برای انجام انتخاب ویژگی در هر گره داخلی درون یک DT چند متغیره استفاده شوند. این الگوریتم ها ویژگی هایی را برای شمول در هر تست بر مبنای داده های مشاهده شده در گره خاصی انتخاب می کنند، و مجموعه یکپارچه ای از ویژگی ها را انتخاب نمی کنند که در آن تست ها مبتنی بر کل درخت باشد.

شکل 2.



علاوه بر این استفاده از الگوریتم های رده بندی مختلف در گره های مختلف یک رده بندی کننده DT ممکن است. این نوع درخت DT پیوند نامیده شده است. مشی دیگر برای طراحی DT استفاده از ماشین های بردار پشتیبان برای طراحی DT با استفاده از ماشین های بردار پشتیبان است که توسط Bennett و Blue پیشنهاد شده اند.

#### 4- نتایج

هدف پژوهش حاضر ارزیابی اثر عوامل زیر روی سطح دقت رده بندی کلی حاصل شده با سه الگوریتم رده بندی منتخب برای این مطالعه است:

اندازه مجموعه داده های آزمایشی

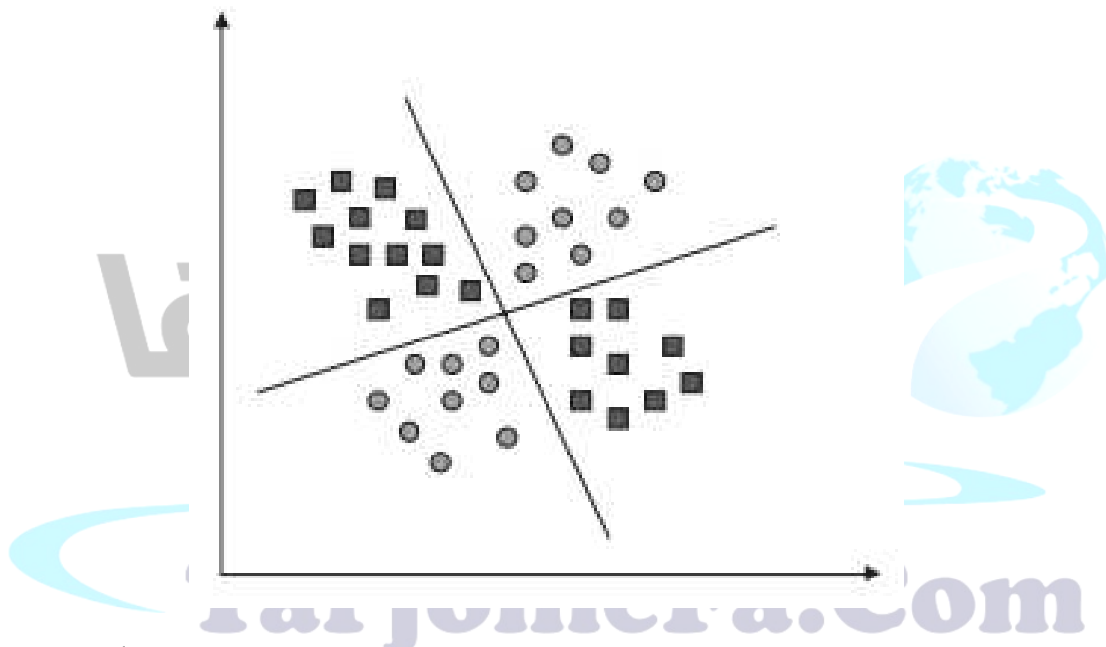
چند بعدی بوده مجموعه داده ها

مقیاس های انتخاب ویژگی

روش های هرس کردن و

تکنیک های تقویت

شکل 3.



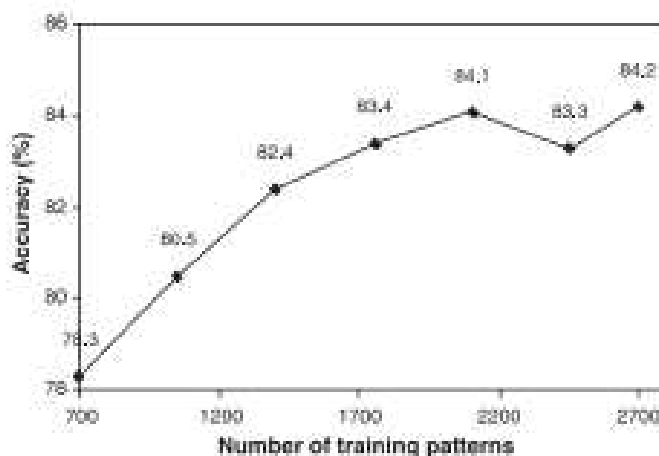
اثر چند بعدی بودن داده ها با استفاده از مجموعه داده های ابر طیفی برای حوزه تست در La Mancha ارزیابی شده اند، در حالی که Landsat EMT+ چند طیفی برای حوزه تست Littleport برای ارزیابی تأثیر تغییرات بر روی دیگر عوامل استفاده شده است. مشخصه های برجسته دو مجموعه داده تست در بخش 2 خلاصه شده اند.

#### 1-4- اثر اندازه مجموعه آموزشی

مشخصه های داده های استفاده شده برای آموزش رده بندی کننده نظارت شده تأثیر قابل توجهی روی دقت رده بندی منتهی دارد. این مهم است که تعداد رده ها برای توصیف پوشش زمین حوزه مطالعه کافی باشد و این که داده های آموزشی شرح مشعری از هر رده ارائه کنند. برای رده بندی کننده ML، الزام مهم این است که تعداد پیکسل های شامل شده در مجموعه داده های آزمایشی برای هر رده باید حداقل 10 تا 30 برابر تعداد ویژگی ها باشد. اندازه

مجموعه آموزشی مورد نیاز می تواند بزرگ باشد، و به سرعت با افزایش تعداد ویژگی ها افزایش می یابد تا از رخداد Hughes جلوگیری شود که نشان دهنده کاهش عملکرد رده بندی کننده با افزایش تعداد ویژگی ها برای یک اندازه مجموعه داده آموزشی ثابت می باشد. بدست آوردن چنین مجموعه های آموزشی بزرگی می تواند مشکل و هزینه بر باشد که در آن تعداد زادی از رده ها شامل شده اند، یا در آن داده های ابر طیفی استفاده شده اند. متعاقباً بررسی هایی وجود دارند که می توانند از اندازه نمونه ای استفاده کنند که کوچکتر از راهنماهای پذیرفته شده برای رده بندی هکنده های آماری نظیر ML است. این اشاره بهاین مطلب دارد که خطاهای استاندارد برای تخمین های پارامترهای مورد نیاز بیشتر از سطح توصیه شده هستند و بنابراین محدودیت های تصمیم می توانند به صورت اشتباه یا با بی دقتی قرار داده شوند. Landgrebe (2000) رابطه بیشتری میان دقت رده بندی و مسأله تعریف کافی از رده در نظر می گیرد.

شکل 4.



اظهار شده که رده بندی کننده های مبتنی بر ANN می توانند با استفاده از مجموعه داده های آزمایشی ای موفق تر عمل کنند که کوچکتر از آن هایی هستند که برای آموزش رده بندی کننده های آماری لازم هستند. با این وجود، بررسی های اثرات مشخصه ای مجموعه آموزشی روی عملکرد ANN ها نشان می دهد که اندازه مجموعه داده های آموزشی اثر مهمی روی دقت رده بندی دارد.

برای ارزیابی اثرات اندازه مجموعه آزمایشی روی دقت رده بندی با استفاده از یک رده بندی کننده DT، هفت زیر مجموعه از داده های آموزشی برای اولین حوزه تست با نمونه گیری تصادفی از مجموعه داده های آموزشی موجود شکل گرفته اند. تعداد پیکسل ها در هر یک از این زیر مجموعه های داده ای آزمایشی به ترتیب 700، 1050، 1400، 1750، 2100، 2400 و 2700 پیکسل هستند که با تعداد مساوی از پیکسل ها در هر رده برای هفت رده با ارائه 100، 150، 200، 250، 300، 350 و 400 پیکسل در هر رده همراه هستند. مجموعه مجزایی از 2037 پیکسل برای تست رده بندی کننده استفاده شده است. مجموعه تست هیچ پیکسلی از مجموعه داده های آموزشی شامل نمی کند.

شکل 4 نشان می دهد که رابطه میان دقت و اندازه مجموعه آزمایشی با استفاده از یک رده بندی کننده DT تک متغیره را نشان می دهد. این نتایج نشان می دهند که سطح دقت با اندازه مجموعه آموزشی افزایش می یابد و این که نرخ افزایش در دقت رده بندی با افزایش اندازه مجموعه آموزشی تا پنجمین مجموعه داده آموزشی خطی است که حاوی 2100 پیکسل می باشد. با افزایش اندازه مجموعه آموزشی از 700 تا 2100 پیکسل، افزایشی در دقت رده بندی حاصل می شود که از  $78/3$  درصد به  $84/1$  درصد می باشد. اما افزایشات بیشتر در اندازه مجموعه آموزشی، با استفاده از مجموعه های داده ای ششم و هفتم نتایجی غیر عادی ایجاد می کنند که با مجموعه داده های آموزشی ششم باعث کاهش اندکی در دقت می شوند.

این نتایج نشان می دهند که (1) دقت یک رده بندی کننده درخت تصمیم تک متغیره با افزایش اندازه مجموعه آموزشی بهبود می یابد، اما فقط تا نقطه مشخصی، و (2) این رده بندی کننده ها برای کارآمد بودن به مجموعه های آموزشی بزرگ نیاز ندارند. باید اظهار کرد که نتایج ما با یافته های Jenson و Qates (1977) مطابقت ندارد که اظهار کردند که اندازه مجموعه داده ای آموزشی هیچ تأثیری روی دقت رده بندی ندارد. یافته های ما نشان می دهند که اندازه مجموعه داده های آموزشی از 300 نمونه در هر رده شرح مناسبی از تغییرات پوشش زمین ارائه می کنند. شکل 300 نمونه خاص مسأله است و نباید از آن به عنوان راهنمایی برای دیگر کاربردها استفاده کرد.

آزمایش دوم برای مطالعه پاسخ یک رده بندی کننده DT چند متغیره برای افزایش اندازه مجموعه داده های آموزشی انجام شد. همان مجموعه داده های آموزشی و تست به صورتی که مثال قبل استفاده شد به کار گرفته شد از سازگاری اطمینان حاصل شود. باز هم دقت رده بندی با افزایش اندازه مجموعه آموزشی افزایش یافت. سپس از آن دقت ش با مجموعه های داده ای پنجم و ششم کاهش یافت، اما باز هم افزایش یافت به گونه ای که بالاترین دقت رده بندی بواسطه مجموعه داده ای هفتم حاصل شد.

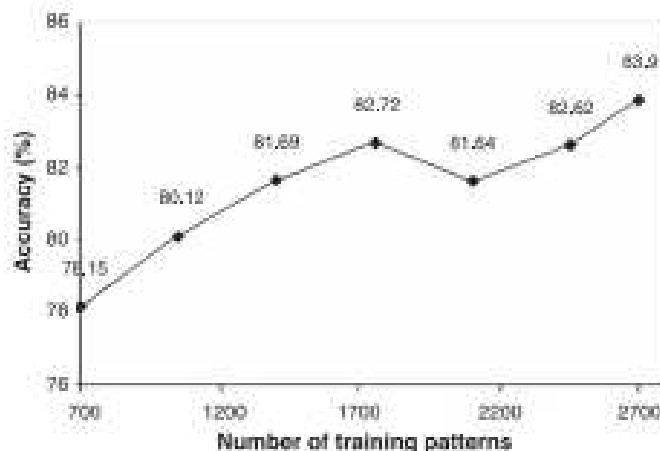
این نتایج نشان می دهند که سطح دقت رده بندی حاصل شده با یک DT چند متغیره با افزایش اندازه مجموعه آزمایشی افزایش می یابد، اما نه به شیوه سیستماتیک. رفتار رده بندی کننده چند متغیره غیر قابل پیش بینی تر از مورد DT تک متغیره بود چون تعداد الگوهای آموزشی فرای حد معینی افزایش می یابند. علاوه بر این واضح است که سطح دقت رده بندی متناظر با رده بندی کننده چند متغیره خیلی بالاتر از سطح دقت رده بندی کننده تک متغیره برای این مجموعه داده ای نیست. چون زمان آموزش همیشه با رده بندی کننده درخت تصمیم چند متغیره بیشتر است، نتیجه می گیریم که رده بندی کننده DT تک متغیره برای این نوع داده مناسب است.

## 2-4- چند بعدی بودن فضای ویژگی

داده های ابر طیفی بواسطه چند بعدی بودن بالایشان توصیف شده اند. مشخصه مهم رده بندی کننده های آماری آن است که ویژگی های هر رده با استفاده از تابع چگالی احتمال مدلسازی شده اند. معمولاً چگالی گاوسی انتخاب شده است، چون می تواند از نظر بردار میانگین و واریانس تعریف شود. خطاهای استاندارد  $k$  عناصر بردار میانگین رده و  $k(k-1)/2$  عناصر ماتریس واریانس - کوواریانس را برای هر رده تخمین می زند و این بسته به نسبت بین تعداد ابعاد، یعنی  $k$ ، و تعداد پیکسل های شامل شده در مجموعه داده های آموزشی برای آن رده است. با افزایش چند بعدی بود داده ها، داده های آموزشی بیشتری برای ارائه تخمین های قابل قبولی از پارامترهای آماری لازم هستند. اگر تعداد پیکسل ها داده های آموزشی کافی نباشد، که می تواند موردی باشد که با داده های ابر طیفی با آن مواجهیم، تخمین پارامترها نادقیق می شود چون خطاهای استاندارد تخمین بیشتر می شود. افزایش تعداد

نوارهای طیفی باعث ارائه اطلاعات بیشتری برای استفاده در تمایز قائل شدن میان رده ها می شود اما برای رده بندی کننده های آماری حداقل این اطلاعات تنها در صورت مفیدند که تعداد داده های آموزشی به نسبت افزایش یابد. هدف این بخش از مطالعه ارزیابی رفتار رده بندی کننده های DT تک متغیره و چند متغیره با افزایش تعداد ویژگی هاست در حالی که اندازه مجموعه داده ای ثابت نگه داشته می شود. چون رده بندی کننده DT از تمامی ویژگی ها به طور همزمان برای آموزش استفاده نمی کنند، و البته چون قابلیت جدا شده رده در داده های با ابعاد بالا می تواند تابعی از تلفی ویژگی ها باشد نه یک ویژگی منفرد، عملکرد رده بندی کننده DT با عملکرد رده بندی کننده های ML و NN مقایسه شده است که از تمام ویژگی های موجود به طور همزمان در فرایند برچسب زنی استفاده می کند. مجموعه آموزش با اندازه ثابت از 2000 پیکسل و مجموعه داده های تست از 3800 پیکسل به کار برده شدند. هر دو از مجموعه داده های تست ابر طیفی استخراج شده بودند.

شکل 5.

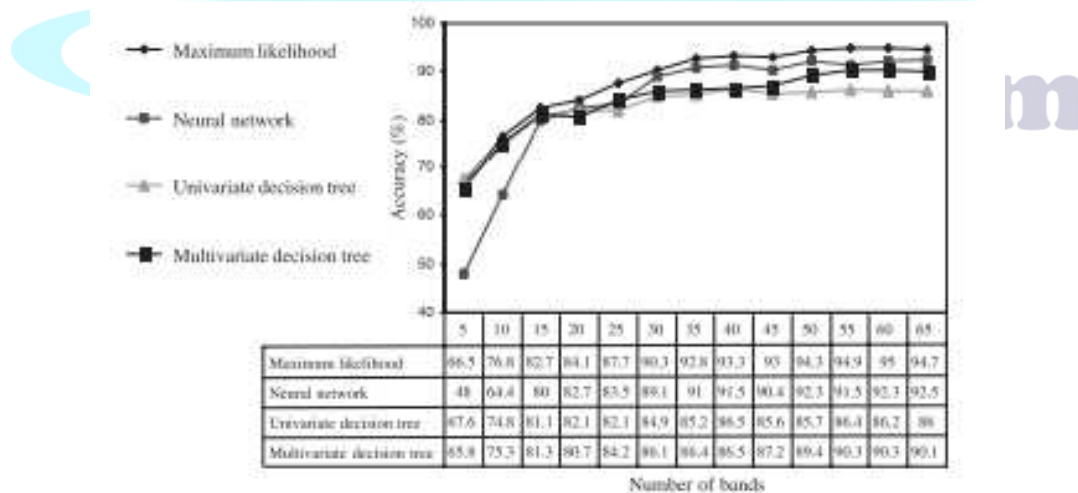


شکل 6 سطوح دقت رده بندی کننده کلی بدست آمده با استفاده از رده بندی کننده های ML، NN و DT را نشان می دهد. تعداد ویژگی ها در ابتدا 5 عدد بود و پس از آن در هر تکرار 5 عدد افزایش داشت، به گونه ای که اولین آزمایش مبتنی بر نوارهای DAIS یک تا پنج، دومین آزمایش مبتنی بر نوارهای 1-10 بود و به همین صورت الی آخر. سطح دقت مرتبط با رده بندی کننده DT تک متغیره بالاتر از مقادیر متناظر برای رده بندی کننده DT، NN، ML و رده بندی کننده های DT چند متغیره برای اولین مجموعه داده ای است که در آن پنج ویژگی

استفاده شده اند اما دقت رده بندی DT تک متغیره با افزایش تعداد ویژگی ها کاهش می یابد. دلیل این رفتار می تواند آن باشد که عملکرد رده بندی کننده DT تک متغیره تحت تأثیر تعداد نمونه های آموزشی است و استفاده از یک نمونه آموزشی بزرگ برای تقسیم بندی فضای ویژگی می تواند حاصل وجود درخت تصمیم بسیار بزرگ و پیچیده باشد. رده بندی کننده DT تک متغیره از تست به کار رفته برای مقدار یک ویژگی منفرد در هر شاخه یا عهر گره در درخت استفاده می کند. با افزایش تعداد ویژگی ها، احتمال این که دو یا چند ویژگی با هم ارتباط پیدا کنند بیشتر می شود. ساختار رده بیشتر به تلفیق هایی از ویژگی وابستگی پیدا می کند و این نتیجه این همبستگی هاست، از این رو عملکرد خوب رده بندی کننده DT تک متغیره مشکل می شود.

بررسی بیشتر با استفاده از DT چند متغیره اظهار دارنده این است که آن به سطح پایین تری از دقت رده بندی با استفاده از داده های چند بعدی بالاتر نسبت به رده بندی کننده ML یا ANN می رسد. پیچیدگی فشاهای ویژگی با ابعاد بیشتر، با ویژگی های مرتبط، می تواند آن قدر کافی باشد تا به رده بندی کننده های DT اجازه بدهد که در مقایسه با رده بندی کننده های ML یا ANN عملکرد بهتری داشته باشند.

شکل 6 .



### 3-4- مقیاس های انتخاب ویژگی

روش های زیادی برای مسأله انتخاب ویژگی های مجموعه ای که باید برای استنتاج DT استفاده شوند وجود دارد، و این روش ها به تفصیل توسط Gebhardt, Borglet, و Kruse (1996), Breiman و همکارانش (1984)،

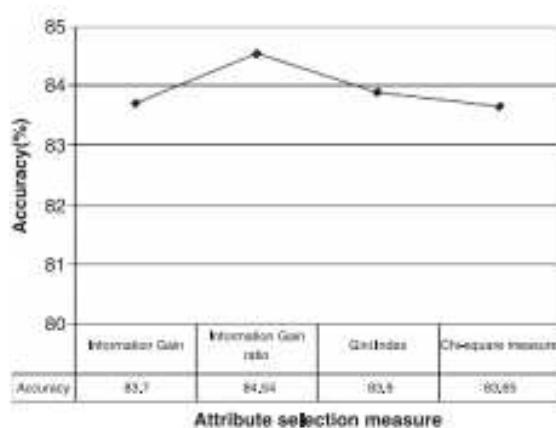


Kononenko و Hong (1997)، Mingers (1989)، Kasif، Murthy و Salzberg (1994) و Quinlan (1993) مورد مطالعه قرار گرفته اند. برخی از این روش های از مقیاس های مناسب بودن تقسیم استفاده می کنند در حالی که دیگر روش ها سعی در به حداقل رساندن ناخالصی داده های آموزشی دارند. یک تابع ناخالصی نایکنواختی یک مجموعه از مشاهدات را می سنجد. این تابع پایین ترین مقدار را برای یک مجموعه خالص و بالاترین مقدار آن را برای یک مجموعه با ناخالصی بیشینه ثبت می کند. توابع ناخالصی در انتخاب ویژگی که باید برای تقسیم بندی بیشتر داده ها در گره فعلی یک DT استفاده شده اند. بهترین ویژگی برای تقسیم بندی با آزمودن این انتخاب شده که هر ویژگی کاندید با چه کیفیتی داده ها را به رده های مختلف تقسیم بندی می کند.

#### 4-4- روش های هرس کردن

رده بندی کننده DT سعی در تقسیم کردن داده های آموزشی به زیر مجموعه هایی دارد که باید تنها حاوی یک رده منفرد باشند. نتیجه این رویه اغلب یم درخت بسیار بزرگ و پیچیده است. در اغلب موارد، متناسب کردن یک DT تا زمانی که تمامی برگ ها حاوی داده ای برای یک رده منفرد باشند می تواند در داده های آموزشی متناسب باشد، چون برخی از نمونه های آموزشی نمی توانند اعضای رده ای شوند که برای نمایش طراحی شده اند. اگر داده های آموزشی حاوی هر گونه خطایی باشند، تناسب بیش از حد داده ها به این شیوه می تواند منجر به عملکرد ضعیف در موارد مشاهده نشده باشد. برای کاهش تأثیر این مشکل، درخت اصلی می تواند هرس شود.

شکل 7.



ساده سازی شامل حذف آن بخش هایی از درخت است که در دقت رده بندی در موارد مشاهده نشده مشارکتی ندارند ، از اینرو درخت هایی با پیچیدگی کمتر و جماع تر تولید می کنند. دو شیوه وجود دارند که در آن ها رده بندی کننده درخت تصمیم می تواند تغییر یابد تا درخت ساده تری ایجاد کند:

- متوقف کردن تقسیم بندی داده های آموزشی قبل از تکمیل داده ها یا
- حذف پشت سر هم بخشی هایی از ساختار درخت با تقسیم بندی بازگشتی.

### جدول 1 .

(a) Effect of pruning on tree size (complexity) and classification error using ETM+ data (test site 1, Littleport)

Evaluation on training data				Evaluation on test data			
Before pruning		After pruning		Before pruning		After pruning	
Tree size	Error (%)	Tree size	Error (%)	Tree size	Error (%)	Tree size	Error (%)
713	1.6	231	8.6	713	17.6	231	15.7

Classification accuracy is defined as  $(100 - \text{error})\%$ . Results indicate that pruning reduces the size of decision tree as well as error on test data. In this case, the size of the pruned tree is  $\sim 33\%$  of the original tree size. The error level of the independent test data set drops by  $\sim 2\%$ .

(b) Classification accuracy from boosted and unboosted decision trees using ETM+ data from test site 1 (Littleport) with a total of 2700 training pixels and seven classes

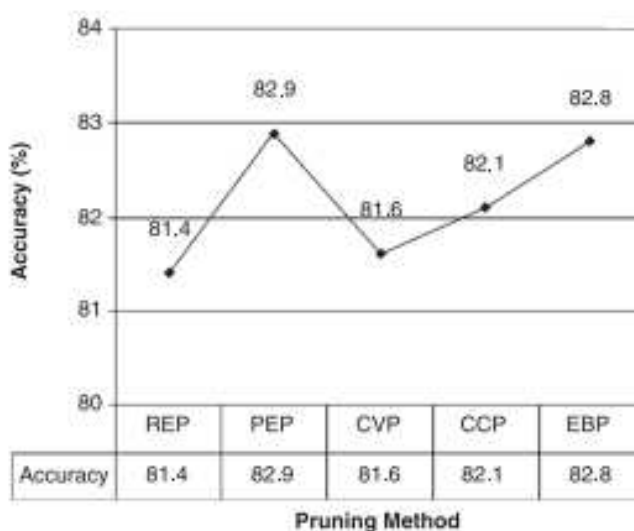
	Accuracy (%)	Kappa value
Unboosted decision tree	84.24	0.816
Boosted decision tree	88.46	0.865

اولین روش، که گاهی اوقات متوقف کردن یا هرس کردن نامیده شده است این مزیت را دارد که در آن زمان در اسمبل کردن ساختاری هدر نمی شود که در درخت ساده شده نهایی مورد استفاده قرار نمی گیرد. این روش به دنبال بهترین روش تقسیم بندی یک مجموعه داده ای بر حسب شرایطی نظیر نفع اطلاعاتی یا کاهش خطاست. اگر مقدار شرط زیر آستانه معینی افت کند، تقسیم بندی بیشتر مجموعه داده ای رد می شود. مشکلی که با این روش داریم تکیه بر فرمولاسیون قانون توقف مناسب دارد. اگر مقدار آستانه بسیار بالا تعیین شده باشد، تقسیم بندی قبل از این که مزایا تقسیمات متعاقب بارز شوند خاتمه یافته است، در حالی که مقدار آستانه بسیار پایین نتیجه ساده سازی کم درخت است.

در روش دوم، به درخت اجازه داده شده تا حوزه کاملش رشد یابد. این درخت شلوغ پس از آن هرس می شود. زمان محاسبه بیشتری برای ساخت آن بخش های از درخت نیاز است که متعاقباً رها شده اند، اما این هزینه در مقابل مزایا منتهی از کاوش کلی تر بخش های ممکن دچار انحراف شده است.

هرس کردن یک DT باعث خواهد شد آن درخت در رده بندی اغلب داده های آموزشی دچار اشتباه شود. از این رو برگ های درخت هرس شده الزاماً حاوی داده های آموزشی یک رده منفرد نمی باشند. در عوض، توزیع رده ای وجود خواهد داشت که برای هر رده احتمال این را مشخص می کند که یک نمونه داده آموزشی در برگ ها متعلق به آن رده هستند. دو خانواده تکنیک برای پیش بینی نرخ خطاهای یک درخت در دسترسند. در خانواده اول، نرخ خطاهای درخت و زیردرخت های آن با استفاده از مجموعه ای از داده های تست پیش بینی شده اند که جدای از داده های آموزشی می باشد. چون این موارد تست در ساخت درخت استفاده شده اند، تخمین دقت رده بندی که از آن ها حاصل شده است پیشقدر خاصی ندارد، و اگر داده های کافی در دسترس باشند، تخمین نیز قابل اعتماد خواهد بود. در روش دوم، خود داده های آموزشی برای پیش بینی این نرخ های خطا استفاده شده اند.

شکل 8 .



این بخش نتایج آزمایشی را توصیف می کند که اثر روش های هرس کردن مختلف را روی دقت رده بندی بررسی می کنند و پنج روش هرس کردن مختلف با نرخ سود اطلاعات به عنوان مقیاس انتخاب ویژگی در یک رده بندی کننده DT تک متغیره استفاده شده اند. روش های هرس کردن به کار گرفته شده: هرس کردن خطای کاهش یافته (ریال \ vnk oxhd ;T ivs) بدبینانه (PEP) و هرس کردن مبتنی بر خطا (EBP) هستند که همگی توسط (Quinlan, 1987, 1993) پیشنهاد شده اند؛ هرس مقدار بحرانی (CVP) پیشنهادی (Mingers, 1989 الف)؛ و

هرس پیچیدگی هزینه (CCP) پیشنهادی Breiman و همکارانش (1984). کل 8 تأثیر روش های هرس کردن مختلف را روی دقت رده بندی نشان می دهد.

## جدول 2 .

Classifier	Accuracy (%)	Kappa value
Maximum likelihood (ML)	82.9	0.80
Neural network (NN)	85.1	0.83

عملکرد روش REP بدترین است، و دقت رده بندی 81/4٪ است. دلیل این می تواند نیاز به مجموعه داده ای مجزایی برای هرس کردن باشد، که نتیجه ای است که Malerba, Esposito و Semeraro (1997) نیز به آن اشاره کرده اند. PEP بالاترین دقت به میزان 82/9٪ را دارد، اما Esposito و همکارانش (1997) اشاره کرده اند که معرفی تصحیح پیوستگی در تخمین نرخ خطا هیچ تعیین نظری ای ندارد و چنین عاملی به اشتباه با یک نرخ خطا مقایسه شده است که می تواند منجر به هرس کم یا هرس بیش از حد درخت شود. عملکرد CVP تحت تأثیر انتخاب مجموعه مقدار بحرانی برای هرس کردن یک درخت است. CCP از مجموعه داده های مجزا یا روش اعتبار متقاطع برای هرس کردن استفاده می کند و زیر درخت هرس شده با به حداقل رساندن عامل پیچیدگی در مجموعه ای از زیر درخت های هرس شده انتخاب شده است. Gelfand, Ravishankar و DElp (1991) اظهار می کنند که این مجموعه زیر درخت ها می توانند ممکن است در زیر درخت بهینه شامل نشود. در نهایت، EBP از داده های آموزشی برای هرس کردن درخت استفاده می کند. این نتایج اظهار دارند این هستند که انتخاب یک روش هرس کردن مناسب عامل مهمی در طراحی رده بندی کننده DT در مقایسه با مقیاس های انتخاب ویژگی می باشد، چون دسترسی پذیری تعداد کافی از عناصر داده های آموزشی همیشه یک مشکل بوده است. ما نتیجه می گیریم که EBP که دقت رده بندی کلی به میزان 82/8 درصد ارائه می کند روش برتر (ترجیحی) ما است. این نتیجه گیری منحصراً مبتنی بر افزایش نسبتاً کم سطح دقت رده بندی نیست که حاصل استفاده از هرس کردن

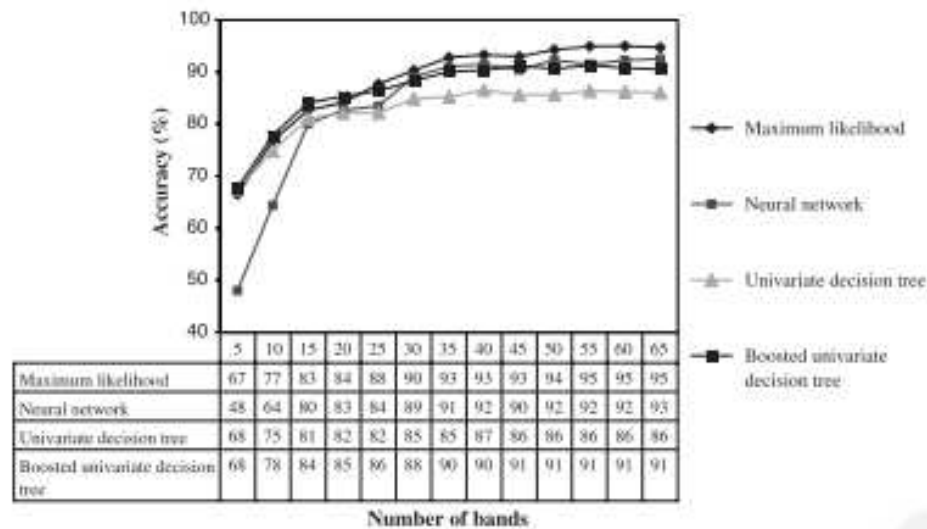
است؛ علاوه بر این مزایای مشتقه از کاهش اندازه درخت و ساده سازی درخت را در نظر می گیرد که می تواند باعث بهبود تفسیر و درک شود.

## 5-4- تقویت

تقویت روش بهبود عملکرد رده بندی کننده ضعیف است. در اصل، این بهبود بواسطه وزن دادن به عناصر فردی مجموعه داده های آموزشی حاصل می شود مقایسه خروجی رده بندی و برچسب معین هر عنصر از داده های آموزشی باید مواردی را نشان دهد که در آن ها عناصر داده های آموزشی به اشتباه رده بندی شده اند. این عناصر داده ای آموزشی که به اشتباه رده بندی شده اند دارای وزن می شوند و رده بندی مجدداً اجرا می شود. وزن افزایش یافته موارد مشکل رده بندی کننده را مجبور می کند تا روی این موارد تمرکز کنند. روش مشابه با تقویت توسط Jackson و Landgrebe (2001) تشریح شده است. ما از الگوریتم AdaBoost M1 همراه با نرم افزار درخت تصمیم C4.5 استفاده کرده ایم.

ترجمه فا

شکل 9.



دقت رده بندی ها و مقادیر کاپای بدست آمده از DT های تقویت نشده و تقویت شده که با استفاده از مجموعه هایی از 2700 داده آموزشی و 2037 داده تست مجزا برای آزمایش ناحیه 1 تخمین زده شده اند در جدول 2 نشان داده شده اند. رده بندی های DT تقویت شده با استفاده از 14 تکرار از الگوریتم درخت تصمیم پایه تخمین زده

شدند. تعداد تکرار های تقویت از 2 تا 20 تکرار متغیر بود اما تغییر اندکی در دقت رده بندی بالای 14 تکرار وجود داشت و میزان بهبود دقت از طریق استفاده از شروع های تقویت تا تثبیت پس از هشت تکرار بدست آمد. مشخص شد که 10 – 15 تکرار تقویت برای بدست آوردن بهبودی در دقت رده بندی ایننوع داده ها کافی است. علاوه بر این نتایج با نتایج مطالعاتی مطابقت دارند که از داده های حسگری استفاده می کردند که از راه دور حاصل نشده بودند. Quinlan (1996) نتیجه می گیرد که حدود 10 تکرار تعداد بهینه است و این که با انجام تقویت های بیشتر بهبود اندکی حاصل می شود. ما به افزایش بیش از 4 درصد در سطح دقت رده بندی پس از تقویت با استفاده از مجموعه داده های تست 1 اشاره کردیم که افزایش کمی در مقایسه با نتایج گزارش شده توسط Quinlan (1996) و Muchoney و همکارانش (2000) می باشد. گرچه سطح بهبود اندک است، باید این را در ذهن داشته باشیم که حتی ایجاد درصد کمی افزایش در بهبود، وقتی سطح دقت رده بندی کلی از 80 درصد بالاتر می رود، کار مشکلی است.

نتایج نشان داده شده در شکل 9 و 12 اظهار دارنده این هستند که تقویت نیز باعث بهبود دقت رده بندی یک DT تک متغیره به میزان 4 درصد می شود و این زمانی است که از آن با داده های ابر طیفی استفاده می شود. بنابراین می توان نتیجه گرفت که بهبود عملکرد یک رده بندی کننده DT برای مطالعات رده بندی پوشش زمین تکنیک مفیدی است.

##### 5- مقایسه درخت تصمیم، ماکسیمم مشابهها (ML) و رده بندی کننده های شبکه عصبی (NN)

در این بخش، نتایج حاصل شده توسط DT تک متغیره در رده بنده داده های تست تشریح یافته در بخش 2 با آن هایی که توسط رده بندی کننده های ML و ANN تولید شده اند مقایسه شده اند. هدف این مقایسه تعیین این مطلب است که آیا مقادیر دقت بالایی که ما با T بدست می آوریم وابسته به رده بندی کننده است یا خیر. ما از رده بندی کننده عصبی پس انتشار استاندارد، با لایه پنهان منفردی استفاده کرده ایم که دارای 26 گره می باشد. تعداد گره ها و مقادیر ورودی هایی که کاربر آن ها در شبکه تعریف کرده است با استفاده از راهنمایی هایی که توسط Kavzoglu (2001) ارائه شده اند تعیین شده اند. جدول 3 دقت های حاصل شده با رده بندی کننده های ML و

NN را با مجموعه داده های تست 1 نشان می دهد که از همان داده های آموزشی و تستی که در بخش های قبل استفاده شده اند استفاده می نماید.

نتایج ارائه شده در جداول 2 و 3 نشان می دهد که رده بندی کننده DT سطح بالاتری از دقت رده بندی نسبت به رده بندی کننده ML تولید می کند، و عملکرد آن قابل قیاس با عملکرد رده بندی کننده ANN است، حتی بدون تقویت. پس از تقویت، سطح دقت رده بندی حاصل شده با DT حدود 3/3 درصد افزایش می یابد. این نمی تواند به دلیل افزایش زیاد باشد، بلکه نشان می دهد که DT تقویت شده تقریباً همان سطح از دقت رده بندی را ارائه می کند که ANN ارائه می کند، در حالی که هم روش ANN و هم روش DT دقت بالاتری نسبت به رده بندی کننده ML ارائه می کنند. اما رده بندی کننده DT تنها مستلزم گزینه انتخاب ویژگی و روش های هرس کردن است در حالی که استفاده از ANN شامل تصمیماتی در رابطه با نوع شبکه، معماری شبکه، و مقادیر اولیه پارامترهای مختلف می باشد. این معمولاً موردی است که در آن هم زمان آموزش مورد نیاز یک رده بندی کننده ANN طولانی است.

جدول 3.

Classifier	Z value
Decision tree (WB) v. Maximum likelihood	2.13
Decision tree (WB) v. Neural network	1.01
Decision tree (B) v. Neural network	2.54

عملکرد درخت های DT تقویت شده و تقویت نشده ای که از داده های ابر طیفی استفاده می کنند با دقت های حاصل شده با استفاده از رده بندی کننده های ML و NN در شکل 9 مقایسه شده است. نتایج این بخش از مطالعه اظهار دارند این است که حتی پس از تقویت، عملکرد رده بندی کننده DT زیر عملکرد رده بندی کننده های ML و NN قرار می گیرد، و این نشان می دهد که عملکرد رده بندی کننده های DT تقویت شده اساساً بسته به عملکرد رده بندی کننده DT مبنا می باشد. نتایج ارائه شده در شکل 9 نشان می دهند که رده بندی کننده ML بالاترین سطح دقت رده بندی را برای مجموعه های داده ای حاوی 20 تا 25 ویژگی یا بیشتر ارائه می کند.

زیر این نقطه، DT تقویت شده تنها اندکی بهتر عمل می کند. برای چک کردن این نتایج، آزمایش با استفاده از نمونه تصادفی متفاوتی از داده های آموزشی با همان اندازه ای که قبلاً استفاده شد تکرار شد. نتایج بسیار نزدیک به نتایجی بودند که توسط آزمایش اول حاصل شدند، و همان استنباط ها پیشنهاد شده اند، یعنی این که در داده هایی که تعداد ابعاد آن ها بیشتر از 20-25 عدد است، هر دو رده بندی کننده ML و ANN دقت های رده بندی کلی بالاتری نسبت به DT تک متغیره تقویت شده یا تقویت نشده تولید می کنند.

هزینه ملاحظه مهمی در کاربردهای عملیاتی حسگری از راه دور است و آموزش یک رده بندی کننده اغلب نشان دهنده نسبت قابل توجه و معناداری از این هزینه هاست. زمان آموزش برای رده بندی کننده ANN حدود 58 دقیقه CPU در ایستگاه کاری دو پردازشگری Sun در مقایسه با 0/7 ثانیه CPU بود که ب از کامپیوتر شخصی با پردازشگر Pentium II fvhd. jr,dj kani hsjthni ld ;vn.

## 6- نتیجه

هدف اصلی این مطالعه ارزیابی سودمندی رده بندی کننده های DT و با سودمندی رده بندی کننده های ANN و ML می باشد. اهداف خاص مطالعه رفتار رده بندی کننده های درخت تصمیم با تغییرات ایجاد شده در اندازه داده های آموزشی، انتخاب مقیاس های انتخاب ویژگی، روش های هرس کردن، و تقویت کردن می باشند. نتایج ارائه بالا اظهار دارنده چندین استنباط می باشند. اولاً عملکرد DT تک متغیره و چند متغیره همیشه متأثر از اندازه مجموعه داده های آموزشی است. این یک خروجی قابل پیش بینی است، اما استفاده از مجموعه های داده ای آموزشی معمول و تست نشان می دهند که رفتار DT تک متغیره سیستماتیک تر از رفتار DT چند متغیره است و این که در مورد DT تک متغیره، حداقل 300 پیکسل در هر رده آموزش برای ارائه مناسب ترین تلفیق از دقت رده بندی و اندازه نمونه لازم است. علاوه بر این مقاله نتیجه می گیرد که رده بندی کننده های DT برای مجموعه داده هایی که ابعاد زیادی دارند مناسب نیستند. نتایج دیگر نشان می دهند که انتخاب یک روش هرس کردن مناسب تأثیر مثبتی در بهبود دقت رده بندی دارد، در حالی که استفاده از مقیاس های انتخاب ویژگی خیلی مهم نبود. استفاده از تقویت



کردن توصیه شده است؛ در این مطالعه، این مورد حاصل دقت رده بندی در حدود 3 - 4٪ با هزینه اندک در زمان کامپیوتر یا پیچیدگی مصرف می باشد.

#### جدول 4 .

Classification accuracies achieved using 2000 training and 3800 test data with all 65 features of the hyperspectral data set (test data set 2, La Mancha)

	Classifier used			
	ML	NN	Univariate decision tree	Boosted univariate decision tree
Accuracy (%)	94.7	92.5	86.0	90.5

مطالعات انجام شده با استفاده از رده بندی کننده های ML و ANN برای همان مجموعه های داده ای نشان می دهند که DT با استفاده از داده های EMT+ اندکی بهتر از ML عمل می کند. عملکرد ANN نیز اندکی بهتر از رده بندی کننده درخت تصمیم تک متغیره تقویت نشده برای داده های EMT+ است، اما تفاوت از نظر آماری معنادار نیست، که این مسأله در جدول 4 نشان داده شده است. تعدادی از مطالعات اظهار دارند این هستند که عملکرد یک رده بندی کننده ANN بسته به مقادیر عددی پارامتریایی است که کاربر باید از قبل تعریف کند، در حالی که عملکرد یک رده بندی کننده DT تک متغیره بسته به روش هرس کردن استفاده شده در طراحی درخت است. زمان آموزش برای رده بندی کننده عصبی در مقایسه با زمان رده بندی کننده درخت تصمیم تک متغیره بیشتر است. حتی با استفاده از تقویت، زمان آموزش برای درخت تصمیم باز هم در مقایسه با نیازهای رده بندی کننده ANN کوتاه است، اما عملکرد درخت تصمیم تقویت شده بهتر از عملکرد رده بندی کننده های ANN است.

وقتی داده های ابر طیفی استفاده شده باشند، عملکرد رده بندی کننده های DT تک متغیره و چند متغیره با افزایش تعداد ویژگی ها افت می کند، در حالی که هم ML و هم ANN مقادیر دقت رده بندی بالاتری نسبت به مقادیر تولید شده توسط رده بندی کننده DT ارائه می کند. این می تواند به دلیل تلفیق الزام اندازه مجموعه داده ای آموزشی بزرگ و استفاده از یک ویژگی منفرد برای تقسیم داده های آموزشی باشد. در نتیجه همبستگی های

میان ویژگی‌هایی که مجموعه داده‌های ابر طیفی را شکل می‌دهند، می‌تواند مورد یباشد که در آن تلفیق از ویژگی‌ها برای اتخاذ تصمیمات آگاهانه لازم است. دلیل احتمالی عملکرد ضعیف رده بندی کننده DT چند متغیره در داده‌هایی که ابعاد زیادی دارند می‌تواند نتیجه انتخاب ویژگی محلی باشد که اظهار دارنده آن است که استفاده از رده بندی کننده های DT با ابعاد بیشتر محدود است. در نهایت نتایج نیز نشان می‌دهند که تقویت یک رده بندی کننده DT تک متغیره عملکرد خوبی برای داده‌های چند بعدی ندارد.

در مجموع، در آزمایشاتی که ما با استفاده از داده‌های EMT+ چند طیفی و داده‌های DAIS ابر طیفی انجام داده ایم، رویه ML عملکرد مناسب یا بهتری از خود نشان داده است. روش DT می‌تواند سطحی از دقت رده بندی ارائه کند که اندکی از سطح دقت رده بندی ML برای داده‌های چند طیفی بالاتر است، اما برای داده‌هایی که تعداد ابعاد آن‌ها بیشتر از 20 تا 25 عدد است، DT تقویت نشده دقتی ارائه می‌کند که بسیار پایین تر است. این نتایج مورد تأیید آزمایش دوم بودند. آن‌ها از این متابعت می‌کند که الگوریتم ML الگوریتم برتر از تسمگر این که دلالتی خاصی برای این باور وجود داشته باشد که داده‌ها از توزیع گاوسی تبعیت نمی‌کنند. همان طور که در بخش 104 اشاره کردیم، این کفایت مجموعه‌های داده‌ای آموزشی و تست در توصیف آسیب پذیر موجود در هر رده احتمالاً نسبت به طبیعت الگوریتم رده بندی مورد استفاده عامل مهمی در تعیین دقت رده بندی می‌باشد، خصوصاً برای داده‌هایی که به وضوح ساختاردهی شده‌اند.

## References

- Bennett, K. P., & Blue, J. A. (1998). A support vector machine approach to decision trees. *Proceedings of the IEEE international joint conference on neural networks, Anchorage, Alaska* (pp. 2396–2401).
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Clarendon Press.
- Borak, J. S., & Strahler, A. H. (1999). Feature selection and land cover classification of a MODIS-like data set for semi-arid environment. *International Journal of Remote Sensing*, 20, 919–938.
- Borgelt, C., Gebhardt, J., & Kruse, R. (1996). Concepts for probabilistic and possibilistic induction of decision trees on real world data. *Proceedings of 4th European congress on intelligent techniques and soft computing, Aachen, Germany, vol. 3* (pp. 1556–1560).
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth.
- Brodley, C. E., & Utgoff, P. E. (1992). *Multivariate versus univariate decision trees*. Technical report 92-8. Department of Computer Science, University of Massachusetts, Amherst, MA, USA.
- Campbell, J. B. (1981). Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering and Remote Sensing*, 47, 355–363.
- Esposito, F., Malerba, D., & Semeraro, G. (1997). A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 476–491.
- Foody, G. M., & Arora, M. K. (1997). An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing*, 18, 799–810.
- Foody, G. M., McCulloch, M. B., & Yates, W. B. (1995). The effects of training set size and composition on artificial neural network. *Photogrammetric Engineering and Remote Sensing*, 58, 1459–1460.
- Freund, Y., & Schapire, R. E. (1996). Experiments with new boosting algorithm. *Proceedings of the thirteenth international conference on*

- machine learning (ICML '96) Bari, Italy, July 3–6, 1996 (pp. 148–156). San Francisco: Morgan Kaufmann.
- Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61, 399–409.
- Gabegan, M., & West, G. (1998). The classification of complex data sets: An operational comparison of artificial neural networks and decision tree classifiers. *Proceedings of the 3rd international conference on geo-computation, University of Bristol, UK, 17–19 September 1998*, available at [http://divcom.otago.ac.nz/SIRC/GeoComp/GeoComp98/61/gc\\_61.htm](http://divcom.otago.ac.nz/SIRC/GeoComp/GeoComp98/61/gc_61.htm), accessed 10 April 2003.
- Gelfand, S. B., Ravishankar, C. S., & Delp, E. J. (1991). An iterative growing and pruning algorithm for classification tree design. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 163–174.
- Hepner, G. F., Logan, T., Ritter, N., & Bryant, N. (1980). Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing*, 56, 469–473.
- Hsieh, P., & Landgrebe, D. (1998). *Classification of high dimensional data*. Technical report—ECE 98-4. School of Electrical and Computer Engineering Purdue University, West Lafayette, IN.
- Hughes, G. F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14, 55–63.
- Jackson, Q., & Landgrebe, S. (2001). An adaptive classifier design for high dimensional data analysis with a limited training data set. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 2664–2679.
- Kavzoglu, T. (2001). *An investigation of the design and use of feed-forward artificial neural networks in the classification of remotely sensed images*. PhD thesis, University of Nottingham, Nottingham, UK.
- Kononenko, I., & Hong, J. S. (1997). Attribute selection for modelling. *Future Generation Computer Systems*, 13, 181–195.
- Landgrebe, D. (2000). On the relationship between class definition precision and classification accuracy in hyperspectral analysis. *International geoscience and remote sensing symposium, Honolulu, Hawaii, 24–28 July 2000*. Available at: [http://www.ece.purdue.edu/~landgreb/IGARSS\\_2000.pdf](http://www.ece.purdue.edu/~landgreb/IGARSS_2000.pdf), accessed 15 August 2002.
- Loh, W. -Y., & Shih, Y. -S. (1997). Split selection methods for classification trees. *Statistica Sinica*, 7, 815–840.
- Mather, P. M. (1999). *Computer processing of remotely-sensed images: An introduction* (2nd ed.). Chichester: Wiley.
- Mingers, J. (1989a). An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, 4, 227–243.
- Mingers, J. (1989b). An empirical comparison of selection measures for decision tree induction. *Machine Learning*, 3, 319–342.
- Muchoney, D., Borak, J., Chi, H., Friedl, M., Gopal, S., Hodges, J., Morrow, N., & Strahler, A. (2000). Application of MODIS global supervised classification model to vegetation and land cover mapping of Central America. *International Journal of Remote Sensing*, 21, 1115–1138.
- Murthy, S. K., Kasif, S., & Salzberg, S. (1994). A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research*, 2, 1–32.
- Oates, T., & Jenson, D. (1997). The effects of training set size on decision tree complexity. *Machine learning, Proceedings of the fourteenth international conference on machine learning* (pp. 254–262). San Francisco, CA: Morgan Kaufmann.
- Paola, J. D., & Schowengerdt, R. A. (1995). A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 981–996.
- Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man–Machine Studies*, 27, 221–234.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo: Morgan Kaufmann.
- Quinlan, J. R. (1996). Bagging, boosting and C4.5. *Thirteenth national conference of artificial intelligence* (pp. 725–730). Portland, OR, USA: American Association for Artificial Intelligence.
- Raudys, S., & Pikelis, V. (1980). On dimensionality, sample size, classification error, and complexity of classification algorithms in pattern recognition. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 3, 242–252.
- Schowengerdt, R. A. (1997). *Remote sensing: Models and methods for image processing*. San Diego: Academic Press.
- Swain, P. H., & Hauska, H. (1977). The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics*, 3, 142–147.
- Utgoff, P. E., & Brodley, C. E. (1990). An incremental method of finding multivariate splits for decision trees. *Machine learning, Proceedings of the seventh international conference on machine learning* (pp. 58–65). Austin, TX: Morgan Kaufmann.
- Wilkinson, G. G. (1997). Open questions in neurocomputing for earth observation. *Neuro-computational in remote sensing data analysis* (pp. 3–13). Berlin: Springer-Verlag.

برای خرید فرمت ورد این ترجمه، بدون واتر مارک، اینجا کلیک نمائید.

این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی