

30 systems for pollen recognition has great potential in all areas of palynology for the accurate
31 and rapid accumulation of data.

32

33 **Keywords:** artificial neural network; automatic identification; expert system; extreme
34 learning machine; *Onopordum*; pollen; Turkey

35

36

37 **1. Introduction and background**

38

39 Pollen grains are produced by seed plants to disseminate their haploid male genetic material.
40 Each pollen grain contains a generative cell (the male gametes) and a vegetative cell or cells,
41 surrounded by a cellulose cell wall and a tough outer wall made of the resistant
42 polysaccharide sporopollenin (Edlund et al. 2004). The morphology of pollen grains is
43 extremely characteristic and pollen can, by itself, be used as a proxy for the respective parent
44 plant. These features are used to identify taxa and hence are useful for establishing
45 phylogenies (e.g. Clark et al. 1980). Pollen analysis is an extremely important discipline and
46 its practitioners, termed palynologists, study diverse topics such as the indications and timings
47 of anthropological activity, limnology, rapid climatic/ecological change and vegetational
48 history (e.g. Moore et al. 1991). Pollen morphology is an essential part of general plant
49 morphology, and hence plays a critical role in research into taxonomy and evolution. Most
50 morphological features of pollen allow identification only to the generic level. This is because
51 the majority of morphological characters are very similar within a genus, and it is normally
52 difficult to subdivide genera using conventional light microscopical techniques.

53 The traditional method of pollen identification using a transmitted light microscope
54 requires an experienced palynologist, and can be somewhat time-consuming. Hence an
55 automated system for the location of pollen grains on microscope slides and their
56 identification would be hugely beneficial in the interests of economics and efficiency in all
57 types of pollen analysis. Several attempts at developing reliable expert systems have been
58 made, and these are reviewed in section 2 below.

59 In this study, an automatic pollen recognition system using a neural network is trialled.
60 A learning algorithm termed the extreme learning machine (ELM) was used to perform the
61 analyses on ten species of a thistle genus *Onopordum* (Family Asteraceae, Subfamily
62 Carduoideae, Tribe Cynareae). The ELM is a single hidden layer feed-forward neural network
63 (SLFN), and is a specialised artificial neural network (ANN) model. With the ELM, the
64 weightings belonging to neurons at the input layer, and the bias values belonging to neurons
65 in the hidden and input layers are all randomly-generated. By contrast, the outputs from the
66 hidden layer are computed analytically (Huang & Siew 2005, Li et al. 2005, Huang et al.
67 2006a,b, Rong et al. 2008, Suresh et al. 2010). The most significant feature of the ELM model
68 is that the learning process is very efficient. It can learn thousands of times faster than
69 conventional learning algorithms for feed-forward neural networks. The learning speed of
70 other feed-forward neural networks is typically relatively slow, largely due to the slow
71 gradient-based learning algorithms used in the training procedure (Huang et al. 2006b).

72 Automated recognition tools such as the ELM, and the necessary computer hardware,
73 are presently at a stage where these methods can potentially be routinely applied to the
74 analysis of pollen assemblages. In theory, automated pollen identification and classification
75 should remove analytical subjectivity and inconsistencies between operators. Furthermore,
76 analyses should be completed more rapidly than with an actual palynologist, hence making
77 savings in terms of both time and labour. Automatic systems can be rapidly programmed to
78 analyse different pollen assemblages in terms of geographical locus, geological age and
79 taxonomic focus (families, genera, species etc.). This makes them potentially more adaptable
80 than any single palynologist.

81

82

283 **2. Previous research on the automated identification of pollen**

84

85 Several studies have attempted the digital identification of pollen using artificial intelligent
86 systems, and selected relevant studies are briefly reviewed here. Early studies include
87 Langford et al. (1990) and Vezey & Skvarla (1990), who undertook research into pollen
88 recognition using the scanning electron microscope (SEM), and achieved promising results.

89 Both these studies developed computer systems which were designed to classify pollen grains
90 based on their surface texture. However SEM analysis is relatively expensive and rather slow,
91 and hence is unsuitable for applications which require data and interpretations in a short
92 timeframe. Benyon et al. (1999) used image analysis to attempt to differentiate eleven
93 allergenic fungal spore genera. This study was based on 24 morphological features extracted
94 from digitised images. These authors found that using linear and quadratic discriminant
95 analysis allowed the recognition of both genera and species with a high level of accuracy.
96 France et al. (2000) developed a new approach to this problem based on improving the quality
97 of the image processing with a traditional optical microscope. These authors were able to
98 differentiate between pollen grains and palynodebris, and to classify three different pollen
99 types correctly. Jones (2000) and Ronneberger (2000) investigated pollen recognition using
100 two-dimensional statistical classification and three-dimensional greyscale invariants with
101 confocal microscopy respectively. Boucher et al. (2002) developed a semi-automatic system
102 for pollen recognition. Digitised three-dimensional photographs of Cupressaceae (cypress),
103 *Olea* (olive), Poaceae (grasses) and Urticaceae (nettle) pollen were image-processed in two
104 and three dimensions, and around 77% of the pollen grains were identified by this system,
105 which worked especially well for pollen from the families Poaceae and Urticaceae.
106 Rodriguez-Damian et al. (2006) developed an automatic system for the identification of
107 species of pollen from the Family Urticaceae using a combination of shape and textural
108 analysis. This system achieved 89% of reliable pollen identifications.

109 Li & Flenley (1999) successfully used texture analysis to identify pollen using
110 transmitted light microscope images with neural network analysis, which is a statistical
111 classifier. Ranzato et al. (2007) developed a microscopic image analysis system. This four-
112 stage process was first used to classify 12 microscopic particle types found in human urine,
113 where it achieved a 93.2% success rate. It was then trained and tested on a set of images of
114 airborne pollen grains, where it generated 83% of positive identifications. Allen et al. (2008)
115 and Holt et al. (2011) developed an automated system that locates, photographs, identifies and
116 counts pollen on a conventional microscope slide. The images in Holt et al. (2011) were
117 analysed with an array of mathematically-defined parameters defined by Zhang et al. (2004),
118 and the feature sets obtained were classified using similar sets from known pollen types. The
119 images produced were then checked by a palynologist. Holt et al. (2011) produced pollen
120 counts which only vary within 1–4% of the results produced conventionally by a palynologist.

121 An innovative methodology to discriminate three species of pollen from the Family
122 Urticaceae (*Parietaria judaica*, *Urtica membranacea* and *Urtica urens*) using computer
123 techniques for the definition of digital shape parameters to represent a pollen grain was
124 developed by de Sá-Otero et al. (2004). This system uses area, diameter, mean distance to
125 centroid and roundness, and achieved an at least 86% success rate.

126 Ticay-Rivas et al. (2011) used Fourier descriptors of the morphological details
127 (geometrical parameters) of 17 honey plant pollen species using discrete cosine transform,
128 together with colour information in order to effect automatic identifications. These authors
129 used a multi-layer neural network, and their method achieved a mean of $96.49\% \pm 1.15$ for
130 successful identifications. Recently Kaya et al. (2013) described an expert computer system
131 using a rough set approach for the automatic classification of 20 types of *Onopordum* pollen.
132 Each pollen grain was comprehensively photographed, with 30 different images captured.
133 Key morphological parameters such as the colpus length, the P/E ratio and the echinae length
134 were measured. The dataset of Kaya et al. (2013) comprised 600 pollen samples; 440 samples
135 were used for training the expert system, and the remaining 160 were used for testing using
136 the rough set approach. This method correctly identified 145 of the 160 pollen grains tested, a
137 success rate of over 90%.

138

139

140 **3. The plant family Asteraceae and the genus *Onopordum***

141

142 This study is an attempt to distinguish species of *Onopordum* L., a genus of thistles within the
143 Family Asteraceae using automatic pollen identification. The Asteraceae are commonly
144 referred to as the aster or daisy family. It is the largest family of flowering plants, and was
145 formerly known as the Compositae (Wagenitz 1976, Bremer 1994, Funk et al. 2005, Panero &
146 Funk 2008). This major plant family is extremely geographically widespread, and is
147 represented by over 1600 genera and approximately 23000 species of herbs, shrubs and trees
148 throughout the world (Kubitzki 2007). Of these taxa, 143 genera and approximately 1484
149 species are present in Turkey (Davis 1975, Özhatay et al. 2009). Pollen grains of the
150 Asteraceae are relatively similar in overall morphology throughout the family. The genus

151 *Onopordum* L. is a thistle genus within the Subfamily Carduoideae of the Asteraceae, and
152 includes around 60 species which inhabit north Africa, west and central Asia, the Canary
153 Islands and Europe (Kubitzki 2007). In Turkey, *Onopordum* comprises 19 species, and 2
154 subspecies (Danin 1975, Davis et al. 1988, Özhatay et al. 1994, Güner et al. 2000).
155 *Onopordum* pollen is oblate-spheroidal in shape and the grains occur as monads (Plate 1).
156 Most of the measurable morphological characters are similar in *Onopordum*, and it is difficult
157 to consistently distinguish the species from one another using normal microscopy techniques.

158

159

160 **4. Material studied**

161

162 The pollen grains of the constituent genera within the Family Asteraceae are morphologically
163 very similar, hence they are eminently suitable for the testing of digital identification
164 methods. Material used in this study was 10 species of *Onopordum* which were collected
165 from wild populations in Turkey. Plant specimens and permanent pollen slides have been
166 deposited in the herbarium and the pollen reference collection respectively of the Department
167 of Biology, Faculty of Science, Yüzüncü Yıl University, 65080 Van, Turkey.

168 Pollen was prepared using the technique of Wodehouse (1959); the mounting medium
169 used was glycerin-jelly mixed with 1% Safranin. The slides were studied using an Olympus
170 CX31 light microscope with a 100x oil immersion objective. Measurements were based on 30
171 images of each of the specimens studied, which were manipulated manually where necessary.
172 The specimens were photographed; the resolution of the digital images was 710×720 pixels.
173 All measurements of the pollen grains were made using Olympus Stream micro-imaging
174 software, a computer program; that automatically calculates the distance from any two points.

175 The polar axis (P) and the equatorial axis (E) were measured in all the specimens, and
176 the P/E ratio calculated. It should be noted that the term equatorial axis is often
177 inappropriately used as a synonym for the equatorial diameter (Punt et al. 2007). Additionally,
178 the colpus length and width, the lengths of the columellae and echinae, and the thicknesses of
179 the exine, intine, nexine and tectum were also measured (Plate 1). These 11 parameters are all
180 used for the identification of pollen grains in the Family Asteraceae, and were deemed to be

181 appropriate for use in digital identification. The pollen terminology of Faegri et al. (1989) and
182 Punt et al. (2007) was used.

183

184

185 **5. The methodology of the extreme learning machine (ELM)**

186

187 Feed-forward neural networks (FFNNs) are ideal classifiers for nonlinear mapping
188 investigations that utilise a gradient descent approach for weights and bias optimisation. The
189 important factors that influence the performance of a traditional FFNN algorithm include
190 three important features. The first are small values for the learning parameters which cause
191 the learning algorithm to converge slowly, whereas higher values lead to instability and
192 divergence to a local minimum. The second is that conventional neural networks may be
193 over-trained using back propagation and normally generate inferior generalisation
194 performance. Finally, gradient descent-based learning is an extremely time consuming
195 process for most applications. To overcome these problems, Huang & Siew (2005), Li et al.
196 (2005) and Huang et al. (2006a,b) proposed a learning algorithm called the extreme learning
197 machine (ELM) for single-hidden layer feed-forward networks (SLFNs). The ELM is a SLFN
198 model in which the input weights are random, and the output weights are obtained
199 analytically (Liang et al. 2006, Yuan et al. 2011). The SLFN structure is illustrated in Figure
200 1. The authors believe that the ELM should be tested in the automatic identification of pollen
201 grains. This method is potentially superior to other methods such as decision tree and linear
202 discriminant analysis. Furthermore, the ELM offers faster learning times than other neural
203 networks. Specifically, the five most important features of the ELM are listed below:

- 204 • The ELM is extremely fast
- 205 • The ELM has better generalisation performance
- 206 • The ELM tends to reach solutions in a straightforward manner without extraneous
207 issues such as local minima, learning rate, momentum rate and over-fitting, which are
208 all encountered in traditional gradient-based learning algorithms

- 209 • The ELM algorithm can be used to train SLFNs, with many non-differentiable
 210 activation functions
- 211 • The ELM randomly chooses and fixes the weights between the input and hidden
 212 neurons based on continuous probability density functions, which is a uniform
 213 distribution function in the range -1 to +1. Then it calculates analytically the weights
 214 between the hidden neurons and the output neurons of the SLFN.

215

216 According to Figure 1, on determining that $X = (X_1, X_2, X_3, \dots, X_N)$ is input and
 217 $Y = (Y_1, Y_2, Y_3, \dots, Y_N)$ is output, the mathematical model with M hidden neurons is defined as
 218 in Suresh et al. (2010):

$$219 \quad \sum_{i=1}^M \beta_i g(W_i X_k + b_i) = O_k \quad , \quad k = 1, 2, 3, \dots, N \quad (1)$$

220 Where $W_i = (W_{i1}, W_{i2}, W_{i3}, \dots, W_{in})$ and $\beta_i = (\beta_{i1}, \beta_{i2}, \beta_{i3}, \dots, \beta_{im})$ are the input and output
 221 weights; b_i is the bias of the hidden neuron and O_k is the output of the network. $g(\cdot)$ denotes
 222 the activation function (Rong et al., 2008).

223 In a network of N training samples, the aim is zero error: $\sum_{k=1}^N (O_k - Y_k) = 0$ or with
 224 minimum error: $\sum_{k=1}^N (O_k - Y_k)^2$. Therefore, Equation 1 can be shown as below (see Huang et
 225 al. 2006b):

226

$$227 \quad \sum_{i=1}^M \beta_i g(W_i X_k + b_i) = Y_k \quad , \quad k = 1, 2, 3, \dots, N \quad (2)$$

228

229 This is because, in the equation above, $g(W_i X_k + b_i)$ denotes the output matrix in the hidden
 230 layer; Equation 2 is therefore as in Huang et al. (2006b):

231

$$232 \quad H\beta = Y \quad (3)$$

233

234 This is where:

$$235 \quad H(W_1, \dots, W_M; b_1, \dots, b_M; X_1, \dots, X_N) = \begin{bmatrix} g(W_1 X_1 + b_1) & \dots & g(W_M X_M + b_M) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ g(W_1 X_N + b_1) & \dots & g(W_M X_N + b_M) \end{bmatrix} \quad (4)$$

236 And

$$237 \quad \beta = \begin{bmatrix} \beta_1^T \\ \cdot \\ \cdot \\ \beta_M^T \end{bmatrix}_{M \times m} \quad \text{and} \quad Y = \begin{bmatrix} Y_1^T \\ \cdot \\ \cdot \\ Y_N^T \end{bmatrix}_{N \times m} \quad (5)$$

238

239 This is where H is the hidden layer output matrix. Training of a network in a traditional feed-
 240 forward ANN means seeking a solution for the least squares in a linear equation of $H\beta = Y$
 241 in the ELM (Suresh et al., 2010).

242 $\hat{\beta} = H^+ Y$ is the smallest norm least-squares of $H\beta = Y$. In addition, H^+ denotes the Moore-
 243 Penrose generalised inverse of the hidden-layer output matrix H . The norm of $\hat{\beta}$ is the
 244 smallest solution among all the least-squares solutions of the $H\beta = Y$ equation (Huang et al.,
 245 2006b). Therefore the ELM can minimise the training error.

246 The ELM algorithm can be summarised in three stages as follows:

- 247 1. The $W_i = (W_{i1}, W_{i2}, W_{i3}, \dots, W_{in})$ input weights and hidden layer b_i bias values
 248 are produced randomly
- 249 2. The H hidden layer output is computed
- 250 3. The $\hat{\beta}$ output weights are computed according to $\hat{\beta} = H^+ Y$. Y is a decision
 251 feature.

252 In this study, an automatic model based on the ELM method was used for the
253 identification of *Onopordum* pollen. A block diagram describing this model is illustrated in
254 Figure 2. The process comprises five blocks, which are summarised below:

255 Block 1: Obtaining 30 images in different orientations for each of the 10 species
256 studied

257 Block 2: Obtaining the key 11 morphometric measurements for each pollen
258 image

259 Block 3: Division of the pollen data sets into training-test partitions at different
260 rates, i.e. 50–50%, 70–30% and 80–20%

261 Block 4: Classification of the training-test partitions through the ELM

262 Block 5: Presentation of the classification results, i.e. the decision stage

263

264

265 **6. Results**

266

267 **6.1. Parameter selection**

268

269 In this study, morphological features that were measured from pollen images were processed
270 by the ELM to effect pollen identification. The 11 parameters used in the ELM network are
271 listed in Table 1. The performance of the ELM network depends on the number of neurons in
272 the hidden layer and the activation function that was used. Consequently, the appropriateness
273 of the parameters in Table 1 were decided as a result of trials. Hence, activation functions
274 such as sigmoid, tangent sigmoid, sine and radial basis were used for the training and testing
275 of the network. The numbers of neurons in the hidden layer between 10 and 100 were
276 finalised by being tested, and this figure was iterated by increasing it one-by-one. The most
277 appropriate activation function and neuron number were finalised only after exhaustive
278 training and testing of the network. For the identification of *Onopordum* pollen, the most
279 appropriate activation function was tangent-sigmoid.

280

281

282 **6.2. Results derived from the experiments using the extreme learning machine**

283

284 The pollen identification experiments were conducted by performing training test sets at the
285 rates of 50–50%, 70–30% and 80–20% through the ELM with the overall pollen dataset. The
286 classification accuracies of these training-test partitions were 84.67%, 91.11% and 95.00%,
287 respectively (Table 2). These accuracies demonstrate that the ELM is consistently very
288 effective. It was found that the ELM has sufficient identification resolution to discriminate
289 *Onopordum* pollen at the species level. In Figures 2, 3, 4 and 5, the ELM performance values
290 related to changes in neuron number used in the hidden layer are illustrated for the training-
291 testing rates of 50–50%, 70–30% and 80–20%, respectively.

292 Different machine learning methods were also used here for automatic pollen
293 identification using the same dataset and images. The accuracies of an artificial neural
294 network (ANN), a support vector machine (SVM; see Chang & Lin 2001), the J48 decision
295 tree method (Quinlan 1993), PART (Eibe & Witten 1998), a logistic regression and the ELM
296 machine learning methods for different training-test partitions were given in Table 2. The
297 ELM gave the highest accuracy for *Onopordum* pollen identification (Table 2).

298

299

300 **7. Conclusions**

301

302 Specific features of pollen can help to identify grains to family or genus level using
303 automated diagnostic systems. These methods potentially allow the accurate and rapid
304 identification of pollen grains, and will be useful in all areas of palynology. In this study,
305 pattern recognition methods were used to determine the pollen type.

306 Morphological characteristics are normally used for identification in plant systematics
307 at all levels from classes to subspecies/varieties. However, at the lower levels, other
308 techniques may be useful to complement the morphological parameters. Pollen morphologies

309 are relatively diverse, and the classification at the family and genus level should be relatively
310 straightforward using traditional microscopy. Computer systems, however, have great
311 potential for performing automatic identifications at the species level and below, due largely
312 to apparent morphological similarities. Hence, the development of automated digital
313 identification systems is predicted to be a significant growth area in the future. The positive
314 results obtained herein from the large and diverse Family Asteraceae, should facilitate more
315 studies on the digital identification of the pollen of other plant families. This field is a rapidly-
316 developing one, and much more experimentation is needed using different characters and
317 criteria in order to improve taxonomic accuracies.

318 In this study, a highly successful approach to automatic pollen recognition and
319 classification using the ELM is demonstrated. The classification process was accomplished
320 using 11 morphological characters for 10 different types of pollen. The identification
321 accuracies of the training-test sections of 50–50%, 70–30% and 80–20% were 84.67%,
322 91.11% and 95.00% respectively (Table 2). The results herein using the ELM compare very
323 well with other expert systems for identifying pollen grains. The identification rate of
324 automatic diagnostic systems will potentially be higher than results obtained manually
325 because of the strict morphometric approach of the former.

326

327

328 **Acknowledgements**

329

330 The authors wish to thank Dr Katherine Holt of Massey University, New Zealand and an
331 anonymous reviewer for their constructive and perceptive critiques of the original manuscript.
332 James B. Riding publishes with the approval of the Executive Director, British Geological
333 Survey (NERC).

334

335

336 **Author biographies**

337

338 YILMAZ KAYA received MSc and PhD degrees in Bioinformatics and Genetics from
339 Yuzuncu Yıl University, Van, Turkey in 2006 and 2011 respectively. Currently he is working
340 in the Department of Computer Science and Engineering at Siirt University in Siirt, Turkey.
341 Yılmaz's research interests include artificial intelligence, machine learning, computer vision
342 systems and statistical modeling.

343

344 S. MESUT PINAR is a botanist working at Yüzüncü Yıl University in Van, Turkey. Mesut
345 specialises on plant taxonomy, and works on a wide variety of domestic research projects. His
346 other research interests include ecology, karyology and palynology.

347

348 M. EMRE EREZ is a plant physiologist working on the relationship between allelopathy and
349 the environment. Emre is a specialist on pollen and seed germination. His other research areas
350 include the use of artificial intelligence in biological systems.

351

352 MEHMET FİDAN is a biologist specialising on systematic botany. He is currently a PhD
353 student working on the plant genus *Gypsophila*. He also researches in molecular polygenies
354 and palynology.

355

356 JAMES B. RIDING is a palynologist with the British Geological Survey based in
357 Nottingham, United Kingdom. Jim is a specialist on Mesozoic-Cenozoic palynology, and
358 works on a wide variety of domestic and international projects. One of his principal tasks is a
359 RCUK Individual Merit research programme entitled *Jurassic dinoflagellate cyst*
360 *palaeobiology and its applications*. His other research interests include forensic palynology,
361 the history of palynology, the palynology of hyperthermal events, palynomorph
362 extraction/preparation and provincialism. Jim is currently the Secretary-Treasurer of the
363 International Federation of Palynological Societies (IFPS).

364

365

366 **References**

367

368 Allen GP, Hodgson RM, Marsland SR, Flenley JR. 2008. Machine vision for automated
369 optical recognition and classification of pollen grains or other singulated microscopic images.
370 Fifteenth International Conference on Mechanotronics and Machine Vision in Practice.
371 Auckland, New Zealand, 2nd–4th December 2008, 221–226.

372

373 Benyon FHL, Jones AS, Tovey ER, Stone G. 1999. Differentiation of allergenic fungal spores
374 by image analysis, with application to aerobiological counts. *Aerobiologia* 15:211–223.

375

376 Boucher A, Hidalgo PJ, Thonnat M, Belmonte J, Galan C, Bonton P, Tomczak R. 2002.
377 Development of a semi-automatic system for pollen recognition. *Aerobiologia* 18:195–201.

378

379 Bremer K. 1994. *Asteraceae: cladistics and classification*. Timber Press, Portland, Oregon,
380 752 p.

381

382 Chang C-C, Lin C-J. 2001. LIBSVM - A Library for Support Vector Machines. URL
383 <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

384

385 Clark WD, Brown GK, Mayes RA. 1980. Pollen morphology of *Haplopappus* and related
386 Genera (Compositae–Astereae). *American Journal of Botany* 67:1391–1393.

387

388 Danin A. 1975. *Onopordum* L. In: Davis PH (ed.), *Flora of Turkey and the East Aegean*
389 *Islands*, Volume 5:356–369, Edinburgh University Press, Edinburgh.

390

391 Davis PH (ed.). 1975. *Flora of Turkey and the East Aegean Islands*, Volume 5. Edinburgh
392 University Press, Edinburgh, 890 p.

393

394 Davis PH, Mill RR, Tan K (eds.). 1988. Flora of Turkey and the East Aegean Islands,
395 Volume 10, Supplement 1. Edinburgh University Press, Edinburgh, 590 p.

396

397 de Sá-Otero MP, González AP, Rodríguez-Damián M, Cernadas E. 2004. Computer-aided
398 identification of allergenic species of Urticaceae pollen. *Grana* 43:224–230.

399

400 Edlund, AF, Swanson, R, Preuss, D. 2004. Pollen and stigma structure and function: the role
401 of diversity in pollination. *The Plant Cell* 16:S84–S97.

402

403 Eibe F, Witten IH. 1998. Generating accurate rule sets without global optimization. In:
404 Shavlik JW (ed.). *ICML '98, Proceedings of the Fifteenth International Conference on*
405 *Machine Learning*, 144–151. Morgan Kaufmann Publishers Incorporated, San Francisco.

406

407 Faegri K, Kaland PE, Krzywinski K. 1989. Textbook of pollen analysis. Fourth edition. John
408 Wiley and Sons, Chichester, 328 p.

409

410 France I, Duller AWG, Duller GAT, Lamb HF. 2000. A new approach to automated pollen
411 analysis. *Quaternary Science Reviews* 19:537–546.

412

413 Funk VA, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling EE, Panero JL,
414 Baldwin BG, García Jacas NT, Susanna A, Jansen RK. 2005. Everywhere but Antarctica:
415 using a supertree to understand the diversity and distribution of the Compositae. *Biologiske*
416 *Skrifter* 55:343–373.

417

418 Güner A, Özhatay N, Ekim T, Başer KHC (eds.). 2000. Flora of Turkey and the East Aegean
419 Islands, Volume 11, Supplement 2. Edinburgh University Press, Edinburgh, 680 p.

- 420
- 421 Holt K, Allen G, Hodgson R, Marsland S, Flenley J. 2011. Progress towards an automated
422 trainable pollen location and classifier system for use in the palynology laboratory. Review of
423 Palynology and Palaeobotany 167:175–183.
- 424
- 425 Huang G-B, Siew C-K. 2005. Extreme learning machine with randomly assigned RBF
426 kernels. International Journal of Information Technology 11:16–24.
- 427
- 428 Huang G-B, Chen L, Siew C-K. 2006a. Universal approximation using incremental
429 constructive feedforward networks with random hidden nodes. IEEE (Institute of Electrical
430 and Electronics Engineers) Transactions on Neural Networks 17:879–892.
- 431
- 432 Huang G-B, Zhu Q-Y, Siew C-K. 2006b. Extreme learning machine: theory and applications.
433 Neurocomputing 70:489–501.
- 434
- 435 Jones AS. 2000. Image analysis applied for aerobiology. Second European Symposium on
436 Aerobiology, Vienna, Austria, p. 2 (abstract).
- 437
- 438 Kaya Y, Pınar SM, Erez ME, Fidan M. 2013. An expert classification system of pollen of
439 *Onopordum* using a rough set approach. Review of Palaeobotany and Palynology 189:50–56.
- 440
- 441 Kubitzki K (ed.). 2007. The families and genera of vascular plants. Volume 9. Flowering
442 plants. Eudicots. Springer-Verlag, Berlin and Heidelberg, 509 p.
- 443
- 444 Langford M, Taylor GE, Flenley JR. 1990. Computerised identification of pollen grains by
445 texture. Review of Palaeobotany and Palynology 64:197–203.
- 446

- 447 Li M-B, Huang G-B, Saratchandran P, Sundararajan N. 2005. Fully complex extreme learning
448 machine. *Neurocomputing* 68:306–314.
- 449
- 450 Li P, Flenley JR. 1999. Pollen texture identification using neural networks *Grana* 38:59–64.
- 451
- 452 Liang N-Y, Saratchandran P, Huang G-B, Sundararajan N. 2006. Classification of mental
453 tasks from EEG signals using extreme learning machine. *International Journal of Neural*
454 *Systems* 16:29–38.
- 455
- 456 Moore PD, Webb JA, Collinson ME. 1991. *Pollen Analysis*. Second Edition. Blackwell
457 Scientific Publications, Oxford, 216 p.
- 458
- 459 Özhatay N, Kültür Ş, Aksoy N. 1994. Check-list of additional taxa to the supplement Flora of
460 Turkey. *Turkish Journal of Botany* 18:497–514.
- 461
- 462 Özhatay N, Kültür Ş, Aslan S. 2009. Check-list of additional taxa to the supplement Flora of
463 Turkey IV. *Turkish Journal of Botany* 33:191–226.
- 464
- 465 Panero JL, Funk VA. 2008. The value of sampling anomalous taxa in phylogenetic studies:
466 major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution* 47:757–782.
- 467
- 468 Punt W, Hoen PP, Blackmore S, Nilsson S, Le Thomas A. 2007. Glossary of pollen and spore
469 terminology. *Review of Palaeobotany and Palynology* 143:1–81.
- 470
- 471 Quinlan JR. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers
472 Incorporated, San Francisco, 302 p.
- 473

- 474 Ranzato M, Taylor PE, House JM, Flagan RC, LeCun Y., Perona P. 2007. Automatic
475 recognition of biological particles in microscopic images. *Pattern Recognition Letters*, 28:31–
476 39.
- 477
- 478 Rodriguez-Damian M, Cernadas E, Formella A, Fernandez-Delgado M, de Sá-Otero MP.
479 2006. Automatic detection and classification of grains of pollen based on shape and texture.
480 *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews*
481 36:531–542.
- 482
- 483 Rong H-J, Ong Y-S, Tan A-H, Zhu Z. 2008. A fast pruned-extreme learning machine for
484 classification problem. *Neurocomputing* 72:359–366.
- 485
- 486 Ronneberger O. 2000. Automated pollen recognition using gray scale invariants on 3D
487 volume image data. *Second European Symposium on Aerobiology, Vienna, Austria*, p. 3
488 (abstract).
- 489
- 490 Suresh S, Saraswathi S, Sundararajan N. 2010. Performance enhancement of extreme learning
491 machine for multi-category sparse data classification problems. *Engineering Applications of*
492 *Artificial Intelligence* 23:1149–1157.
- 493
- 494 Ticay-Rivas JM, del Pozo-Baños M, Travieso CM, Arroyo-Hernández J, Pérez ST, Alonso
495 JB, Mora-Mora F. 2011. Pollen classification based on geometrical, descriptors and colour
496 features using decorrelation stretching method. In: Iliadis, LS, Maglogiannis, I, Papadopoulos
497 H. (eds.), *IFIP Advances in Information and Communication Technology* 364:342–349.
- 498
- 499 Vezey EL, Skvarla, JJ. 1990. Computerized feature analysis of exine sculpture patterns.
500 *Review of Palaeobotany and Palynology* 64:187–196.
- 501

502 Wagenitz G. 1976. Systematics and phylogeny of the Compositae (Asteraceae). *Plant*
 503 *Systematics and Evolution* 125:29–46.

504

505 Wodehouse RP. 1959. *Pollen grains: their structure, identification, and significance in science*
 506 *and medicine*. Hafner Publishing Company, New York, 574 p.

507

508 Yuan Q, Weidong Z, Shufang L, Dongmei C. 2011. Epileptic EEG classification based on
 509 Extreme learning machine and nonlinear features. *Epilepsy Research* 96:29–38.

510

511 Zhang Y, Fountain DW, Hodgson RM, Flenley JR, Gunetileke S. 2004. Towards automation
 512 of palynology 3: pollen pattern recognition using Gabor transforms and digital moments.
 513 *Journal of Quaternary Science* 19:763–768.

514

515

Morphological feature/parameter	Definition
P	The length of the polar axis
E	The length of the equatorial axis
P/E	The P/E ratio
Colpus (L)	The length of the colpus
Colpus (W)	The width of the colpus
Exine	The thickness of the exine
Intine	The thickness of the intine
Nexine	The thickness of the nexine
Tectine	The thickness of the tectine
Echinae	The length of the echinae
Columellae	The length of the columellae

516

517 Table 1. The 11 training parameters (morphological features) used with the extreme
 518 learning machine (ELM) network in this study.

519

520

Name of automatic system	50–50% training-test (%)	70–30% training-test (%)	80–20% training-test (%)
Artificial Neural Network	80.00	80.66	84.44
Extreme Learning Machine	84.67	91.11	95.00
J48 Decision Tree	72.00	81.11	85.00
Logistic Regression	68.88	76.00	76.66
PART	75.33	75.55	83.33
Support Vector Machine	78.66	86.66	88.33

521

522 Table 2. The performance values of automatic pollen identifications using six different
523 automatic systems. The extreme learning machine (ELM) results are in bold font.

524

525 **Display material captions:**

526

527 Figure 1. The structure of a single-hidden layer feed-forward (SLFN) artificial neural
528 network.

529

530 Figure 2. A block diagram illustrating the method for pollen identification used herein.

531

532 Figure 3. Training and test efficiencies for the 50–50% training-test partition.

533

534 Figure 4. Training and test efficiencies for the 70–30% training-test partition.

535

536 Figure 5. Training and test efficiencies for the 80–20% training-test partition.

537

538 Plate 1. Two images of *Onopordum* pollen illustrating the various morphological
539 measurements made in this study. 1 – grain in polar view. 2 – grain in equatorial/lateral view.