J Digit Imaging (2017) 30:640–647
DOI 10.1007/s10278-017-0001-7

Bone Tumor Diagnosis Using a Naïve Bayesian Model of Demographic and Radiographic Features

Bao H. Do[1] · **Curtis Langlotz**[2] · **Christopher F. Beaulieu**[2]

Published online: 27 July 2017

Abstract Because many bone tumors have a variety of appearances and are uncommon, few radiologists develop sufficient expertise to guide optimal management. Bayesian inference can guide decision-making by computing probabilities of multiple diagnoses to generate a differential. We built and validated a naïve Bayes machine (NBM) that processes 18 demographic and radiographic features. We reviewed over 1664 analog radiographic cases of bone tumors and selected 811 cases (66 diagnoses) for annotation using a quantitative imaging platform. Leave-one-out cross validation was performed. Primary accuracy was defined as the correct pathological diagnosis as the top machine prediction. Differential accuracy was defined as whether the correct pathological diagnosis was within the top three predictions. For the 29 most common diagnoses (710 cases), primary accuracy was 44%, and differential accuracy was 60%. For the top 10 most common diagnoses (478 cases), primary accuracy was 62%, and differential accuracy was 80%. The machine returned relevant diagnoses for the majority of unknown test cases and may be a feasible alternative to machine learning approaches such as deep neural networks or support vector machines that typically require larger training data (our model required a minimum of five samples per diagnosis) and are "black boxes" (our model can provide details of probability calculations to identify features that most significantly contribute to truth diagnoses). Finally, our Bayes model was designed to scale and "learn" from external data, enabling incorporation of outside knowledge such as Dahlin's *Bone Tumors*, a reference of anatomic and demographic statistics of more than 10,000 tumors.

Keywords Naïve Bayes model · Bone tumor diagnosis

Introduction

Radiographic interpretation of bone tumors requires identification and processing of multiple demographic and observational features that may correlate with a diagnosis, such as age, sex, tumor margin, matrix, and location [1, 2]. Benign and malignant bone tumors have a wide variety of appearances. Because many bone lesions are uncommon or rare, few radiologists develop sufficient expertise to diagnose bone lesions accurately. In clinical practice, one relies on learning characteristic imaging features of various lesions and recall, both of which are subject to bias. Thus, among general radiologists, interpretation of bone lesions can be variable, leading to misdiagnosis and suboptimal patient management.

Probabilistic approaches to medical diagnosis have been implemented in many areas of medicine, including radiology [3, 4]. Over 50 years ago, Lodwick provided initial proof that computing odds based on structured annotation of bone lesions could improve diagnosis, but the limitations of analog radiography and the lack of rapid computing technology restricted the impact of his work [5, 6].

Bayesian inference can guide decision-making by computing the conditional probabilities of multiple diagnoses/classes (class posterior probabilities) based on observations (attributes/features) and their likelihood (prior probability).

$$P(diagnosis_i | findings) \propto P(findings \mid diagnosis_i) * P(diagnosis_i)$$

✉ Bao H. Do
baodo@stanford.edu

[1] Department of Radiology, VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304, USA

[2] Department of Radiology, Stanford University, 300 Pasteur Drive, Stanford, CA 94305, USA

The conditional probabilities of the array of conditional probabilities can then be sorted to generate a "differential diagnosis."

Alternative quantitative or machine learning approaches have been investigated such as artificial neural networks or support vector machines, but these algorithms provide more abstract "confidence scores" [7]. In 2001, Kahn et al. developed OncOs, a Bayesian network for classification of 10 bone tumor diagnoses [8]. In 2005, based on Lodwick's work, Richardson developed an online Bayesian-based bone tumor decision support tool that incorporated five features (age, size in cm, bone type, longitudinal location, and matrix) to classify nine tumor diagnoses [9]. Preliminary results from these prior works are promising; however, these previous systems were limited to 10 tumor diagnoses. The World Health Organization (WHO) defines at least 20+ primary musculoskeletal osseous tumors [10].

Stimulated by the work of Lodwick and others, we followed a comprehensive, structured, and standards-based approach for construction and evaluation of a naïve Bayes model (NBM) for generating differential diagnosis of bone tumors. Based on a historical teaching collection that started with 1664 cases of more than 60 types of primary and secondary bone tumors and associated syndromes, we used ePad, a web-based structured annotation tool, to annotate 811 cases with 18 semantic features each [11]. These 811 images and their associated annotations represent the structured data used to create the NBM.

Materials and Methods

Institutional review board approval was obtained. The requirement for informed consent was waived as this was a retrospective review of de-identified radiologic images with only age, gender, and brief clinical notes/diagnosis.

Case Selection

The raw data set is a collection of 1664 analog radiographic cases of bone tumors at a tertiary care teaching hospital. Cases were collected by one professor (none of the authors in this work are owners of the collection) between approximately 1955 and 2005 and in almost all instances were copy films. Two students used a transparency film scanner (PACSGEAR—Lexmark, Pleasanton, CA) to digitize all radiographs in each case at 600 dpi. A total of 22,864 images were captured from the 1664 cases. Upon review by a musculoskeletal radiologist, cases were subjectively categorized into 124 low-quality, 675 medium-quality, and 865 high-quality cases. High-quality cases included excellent representation of the lesion in terms of radiographic exposure and resolution, as well as lack of extraneous markings such as wax pencil or film labels. Low-quality cases included under- or over-exposed images that may have exhibited motion artifact or interfering overlying markings. Taking the high-quality cases and a selection of the medium-quality cases, a "top 1000" collection was constructed which included all of the relevant radiographic projections for each lesion; 2147 separate images comprise this collection. During the annotation process (see below), 189 cases were not annotated fully because they had limited visibility of the lesions (more common for lesions in the spine or facial bones, for example) or subjectively lower overall image quality. This curation process resulted in 811 cases. The pathologic diagnosis was obtained by histology for the majority of cases, with a minority of cases diagnosed by pathognomonic features and imaging follow-up, resulting in 66 unique diagnoses.

Feature Selection

The naive Bayes model contains 18 input attributes/features (2 clinical and 16 qualitative radiographic features) (Fig. 1a, b). Feature selection was based on clinical experience and existing knowledge of radiologic observations commonly used to characterize bone tumors, such as bony expansion, bone location (transverse and axial), and patient age (in decades). Bone tumor literature was reviewed to refine or add features that show either high acceptance by the community or statistical significance in stratifying disease, such as border type (1ABC, 2, 3) and endosteal scalloping [1, 12–16]. For example, we initially considered using an attribute called "cortical destruction" with assigned binary values: yes/no. Based on the work by Murphey et al. on x-ray cortical observations in chondroid tumors, we adopted their more formal and quantitative grading system "depth of endosteal scalloping" which ranks endosteal invasion by depth, from 0 (cortex normal) to 4 (full-thickness cortical breakthrough/destruction) [12].

Annotation

For each of the 811 cases, 1 image (AP view or image with best visualization of the tumor, excluding film markings, film degradation, or overlapping bowel/bones/landmarks if possible) was selected for annotation with the 18 clinical and qualitative features by MSK radiologists (XX, 22 years experience, and YY, 6 years experience). All annotations were reviewed for discrepancies by XX who served as adjudicator and determined a consensus. Interobserver agreement statistics were not provided by ePad and were not measured. For consistent encoding, annotations were performed using ePad, a freely available quantitative imaging informatics platform which provides an implementation of the Annotation and Image Markup (AIM) standard [11]. Attribute values for each field were derived from RadLex if possible [17], and

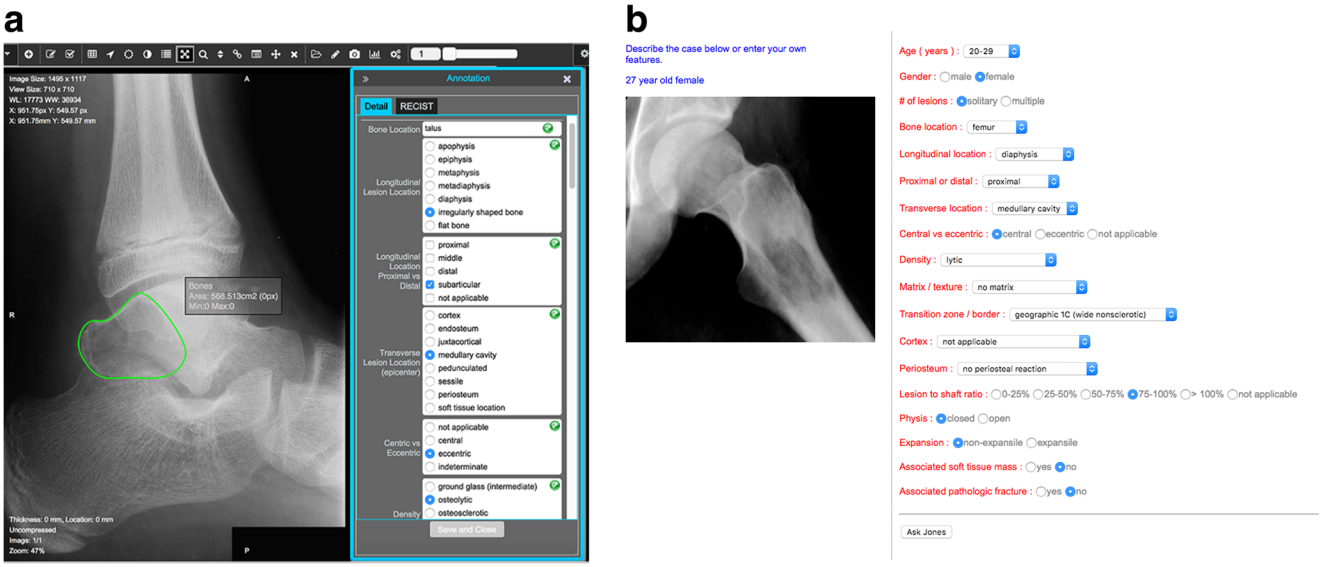

Fig. 1 **a** ePad annotation interface. The image is displayed, and the region of interest is drawn (*green outline*). The annotation template is filled in by clicking radio buttons on the interface. The program requires completion of all fields before annotation is validated and saved. This lesion is an aneurysmal bone cyst of the talus. **b** More complete display of the 18 fields annotated for each case (color figure online)

supplemented by attributes derived from clinical experience or literature descriptions of bone tumor observations [12].

Validation

While a total of 811 fully annotated cases/images were available, several diagnoses had small sample sizes constituting only one to three examples, such as Erdheim-Chester disease and Rosai-Dorfman. To avoid over-fitting, these sparse diagnoses were excluded from further study. By including only diseases that were represented by at least five case examples, the overall experimental dataset was reduced to 710 cases. To assess the effect of sample size on algorithm performance, we created three subsets from the collection of 710 cases using a progressively higher cutoff for minimum sample size, as follows:

Subset A (Most Inclusive, Includes Common, Intermediate, and Rare Cases)

Minimum five examples of each diagnosis: 710 cases, 29 unique diagnoses, shown in Table 1.

* Note that "rare" reflects the relative number of examples of the diagnosis within our population, and not reflective of disease prevalence. Subset A uses a cutoff of at least five samples per diagnosis and therefore includes "rare," intermediate, and common prevalence diagnosis.

Subset B (Includes Intermediate and Rare Cases)

Minimum 18 examples of each diagnosis: 559 cases, 14 unique diagnoses.

Subset C (Includes Common Diagnosis Only)

Minimum 30 examples of each diagnosis: 478 cases, 10 unique diagnoses.

For each subset A, B, and C, a leave-one-out cross validation analysis was performed. For each test input, the machine produced a ranked list of diagnoses with each diagnosis assigned a probability. We established two outcome metrics by comparing machine outputs with the proven diagnosis for each test case. Specifically, we defined *primary accuracy* as whether the correct pathological diagnosis was the top-ranked machine output; we defined *differential accuracy* as whether the correct pathological diagnosis was within the top three machine predictions. The latter metric is intended to reflect clinical practice by providing a short differential list of possible diagnoses.

Bayes Model

The naive Bayes model (NBM) was developed using modern web technologies (front-end JavaScript/HTML5; backend and database PHP/MySQL running on an APACHE web server). The system is fully operational but is used only for research. The classifier generates probabilities for all diagnoses based on the 18 clinical and observational attributes (Table 3), which are calculated in real time from knowledge encoded in the database. Conditional probabilities for all diagnoses are computed and sorted in descending order. The diagnoses are displayed in descending probability as a ranked "differential diagnosis" (Fig. 2). Table 4 shows the raw computations that are generated by the NBM in real time for any unknown query.

Table 1 Bone tumor diagnosis/classes (29 total) represented in the final 710 cases used to test the Bayes machine, ordered from most to least samples

Bone tumor diagnosis (no. of samples)
Group 1
Osteosarcoma (83)
Enchondroma (65)
Metastasis (55)
Osteochondroma (46)
Aneurysmal bone cyst (41)
Chondrosarcoma (41)
Giant cell tumor (41)
Non-ossifying fibroma (38)
Ewing sarcoma (38)
Fibrous dysplasia (30)
Group 2
Lymphoma (22)
Chondroblastoma (21)
Simple bone cyst (19)
Eosinophilic granuloma (19)
Group 3
Osteoid osteoma (16)
Non-Hodgkin lymphoma (16)
Malignant fibrous histiocytoma (14)
Chondromyxoid fibroma (11)
Osteomyelitis (11)
Periosteal chondroma (10)
Multiple myeloma (10)
Osteoblastoma (10)
Ganglion cyst (9)
Paget disease (9)
Giant cell reparative granuloma (8)
Hemangioma (7)
Intraosseous lipoma (7)
Adamantinoma (7)
Plasmacytoma (6)

In the leave-one-out validations, subset A consists of all 29 diagnoses (groups 1, 2, and 3) and is most inclusive; subset B consists of 14 diagnoses (groups 1 and 2); subset C consists of 10 diagnoses (group 1)

Results

The results of the cross validation studies are summarized in Table 2. For subset A (710 cases, 29 diagnoses), primary accuracy was 44%, and differential accuracy was 60%. In other words, the machine predicted the correct diagnosis as the top-ranked result in 44% of leave-one-out trials, and predicted the correct diagnosis with the top three ranked results in 60% of the trials. For subset B (559 cases, 14 diagnoses), primary accuracy was 56%, and differential accuracy was 73%. For subset C (478 cases, 10 diagnoses), primary accuracy was 62%, and differential accuracy was 80%.

Discussion

We constructed and evaluated a naïve Bayes model for predicting bone tumor diagnosis on a collection of more than 60 unique primary and secondary bone tumors and tumor-like conditions derived from a historical collection that began with 1664 digitized radiographic cases. To avoid over-fitting, we restricted the final dataset to 710 cases so that each diagnosis was represented by a minimum of five samples. The strength of a Bayesian model is that relative probabilities can be provided along with diagnosis predictions, allowing for a likelihood ranked "differential diagnosis" (Fig. 2), similar to what is typically provided in the radiology report. An accuracy metric for the differential diagnosis makes both clinical and intuitive sense. Alternative classifiers (e.g., support vector machine, k-nearest neighbor, random forest) cannot provide a ranked differential diagnoses based on probabilities, but instead, more abstract "confidence" scores or relative ranks [7].

In Lodwick's original report, 77.9% of cases were correctly diagnosed as one of eight tumor types in a dataset of 77 tumors [5]. Kahn and co-workers assessed their Bayesian network by having medical students encode 28 cases comprising 10 diagnoses and feeding the features into their model. In that experiment, the correct diagnosis was obtained as the top choice in 68% of cases and within the top two results in 89%. In our work, restricting to 10 diagnoses was achieved by requiring a minimum of 30 examples of each diagnosis (subset C; Table 1). In this subset of 478 cases, our primary and differential accuracy was 62 and 80%, respectively. These results are comparable to those reported by Lodwick and Kahn, particularly given many differences in experimental details, including total case number, features encoded, and computational approaches. Our work also incorporated a much more diverse dataset comprising 29 diagnoses.

Our work demonstrates that Bayesian models can provide meaningful results for small training sets, although clinical utility was not tested in this work. This work focused on development and preliminary validation using a robust leave-one-out technique for all samples in the data. Despite "low" primary and differential accuracy, these accuracies do not reflect *clinical utility*. For example, theoretically, if a system is only 60% accurate (primary diagnosis or prediction) but more correct compared to radiologists 90% of the time, it is potentially *clinically useful*.

The NBM provided a differential accuracy of 60% among 29 unique diagnoses and required only five samples per diagnosis for training/knowledge. By comparison, a deep neural network may require thousands to train a diagnosis [18–21]. Though our results with small sample sizes are encouraging,

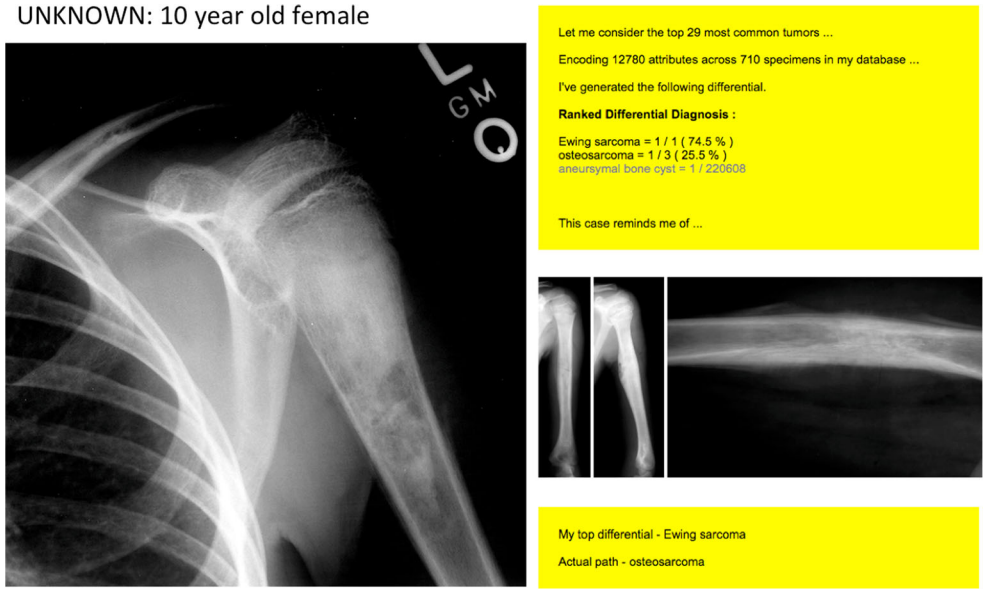

Fig. 2 Example machine output. The machine generates probabilities for all diagnoses based on the 18 clinical and observational instances of an unknown (*left image*). The diagnoses with conditional probabilities >0% are displayed in descending order as a "Ranked Differential Diagnosis" (*upper right*, *yellow box*). Raw fractions are displayed for probabilities <1%. The *differential accuracy* is whether the correct pathological diagnosis is within the top three machine predictions. After ranking the diagnoses, the system displays images from cases with the same top differential and bone location (*middle right column images*). The highest ranked diagnosis or top differential is compared with truth (*lower right yellow box*)

we did demonstrate improved accuracy by increasing the sample size from five in subset A (primary accuracy 44%) to 30 in subset C (primary accuracy 62%). However, more sample training size does not always result in improved accuracy of classification. In a two-layer neural network for bone lesions by Reinus, 29% (4/14) of diagnoses with training size of at least 20 had among the lowest primary accuracy scores (13–39%), and overall performance improved only marginally for diagnosis trained with more samples [7]. Despite this performance, Dr. Reinus' and similar work are novel examples of the promise of unsupervised machine learning in musculoskeletal radiology.

Unlike some radiology classification tasks in which accuracy is based on a binary decision between "benign" and "malignant," the wide variety of possible diagnoses in this data set adds complexity to the diagnostic task. There is a trade-off between clinical utility for decision support (more diagnoses) and accuracy (reduced using fewer training samples). Therefore, the NBM is designed for continuous "learning" as the Bayes probabilities are calculated in real time, and incremental "knowledge" or new labeled samples can be added to the system any time. Bone tumors are also unique in that external data is available, such as Dahlin's textbook reporting the skeletal distribution of lesions as well as patient gender and age (22 cases). These probabilities reflecting disease incidence can be useful adjuncts to the features we have encoded for our specific dataset.

Table 2 Bayes machine performance showing primary and differential accuracy using leave-one-out validation

Subset	Minimum no. of samples/ diagnosis cutoff	Total cases	Distinct diagnosis	Naïve Bayes primary accuracy	Naïve Bayes differential accuracy
A	5	710	29	44%	60%
B	18	559	14	56%	73%
C	30	478	10	62%	80%

Three subsets were trained and validated by varying the size of the minimum cutoff number for class training

An important distinction should be made between retrieval and ranking. Even if image retrieval systems can find a labeled sample identical in appearance to the unknown query, the actual diagnosis can be different from the retrieved image because disease can have overlapping imaging appearances. Ideal intelligent machines *should* suggest a diagnosis based on similar appearances, but, more importantly, provide a differential of statistically likely alternates based on demographic, clinical, and observational data (Fig. 3). Bayesian models hold a significant advantage by accounting for disease prevalence. Alternative machine learning approaches such as deep learning or support vector machines provide abstract "confidence scores" [7] and are "black boxes." Table 4 shows example raw computations that are generated by the NBM in real time for any unknown query. Although this output is normally suppressed to the user, our NBM can display this in debug

mode, allowing for the user to understand the machine's decision process, specifically identifying the attribute/value vectors that most significantly contribute to each diagnoses.

We consider the results reported here to be preliminary, as the outcomes will likely change with increasing case numbers, refinements of features annotated, and improvements in the computational model. The current image dataset itself is limited, as it represents a set of cases collected over many years based on the interests of a single radiologist, with varying image quality. Furthermore, the prevalence of cases within the dataset does not reflect the day-to-day prevalence of focal bone lesions presenting for workup. There may also be qualitative differences in feature selection between the two musculoskeletal radiologists. The current work also is limited because it is based entirely on semantic descriptions of each lesion's visual appearance. The use of image processing and computer vision techniques to directly analyze edge, shape, size, and texture features of bone lesions on radiographs is a promising alternative [22].

There are additional limitations of our work inherent to Bayesian algorithms. Bayes classifiers assume that features are independent (naïve), and that each feature contributes independently to class probability. Common pre-processing steps to identify strong correlations that can lead to overconfidence include computing a weighted correlation matrix for continuous variables or performing chi-square tests for nominal variable pairs. A second limitation of our model is the assumption that attribute values exist for all examples in the training set, and if they do not, then the differential should not be considered. For example, if osteosarcoma always exhibits osteoid matrix, and if an unknown exhibits chondroid matrix, then the class probability for osteosarcoma of the unknown instance is zeroed out (the class probability is the product of each of the posteriors). Laplace correction is a technique used in Bayesian models to correct for the situation in which an attribute value has never before been seen in the population by inserting small default probabilities for missing attributes instead of zero, enabling probability calculation for sparse data [23]. These implementations and their validation were beyond the scope of this work.

The eventual goal of a machine learning system is to improve clinical diagnosis, and ultimately patient outcomes. Neither of these goals has been tested with our NBM. Indeed, we could find no useful reference data on the actual clinical accuracy of bone tumor diagnosis as currently performed by radiologists without computer support. Lodwick estimated that he was 80% accurate in predicting the histologic subtype of bone tumors, but his accuracy varied across tumor types [5]. Based on our own intuition, an accuracy of 80% is probably reasonable for highly experienced experts. Tumor appearances overlap considerably, and histological analysis is ultimately needed for final diagnosis, even though pathologists cannot always agree on bone tumor diagnoses.

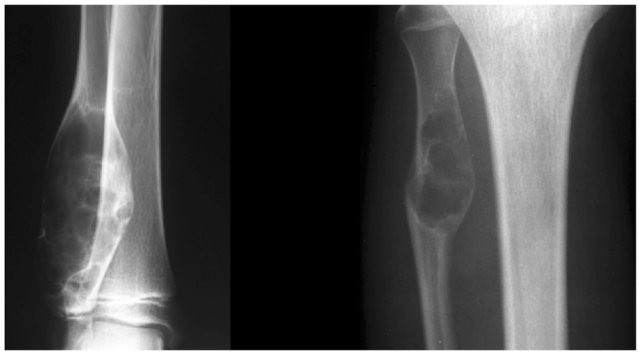

Fig. 3 Two expansile bone lesions with the same imaging attributes but of different tumors: aneurysmal bone cyst (*left*) and non-ossifying fibroma (*right*)

No individual is likely to achieve 100% accuracy. Thus, the output of a machine learning system should be compared against the best possible human performance. Looking toward future work, one of our goals is to test the ability of our system to improve diagnostic accuracy across readers with a wide range of experience, with the hypothesis that decision support will benefit those with less experience [22].

Conclusions

We built and implemented a probabilistic classifier based on a naive Bayesian model that incorporates 18 features, both radiographic observations and demographic characteristics, to rank bone tumor diagnoses. During these trials, the model returned relevant diagnoses for the majority of unknown test cases. Using this approach, we were able to classify a wide range (29 types) of lesion diagnoses, with the potential to classify more than 60 diagnoses if we can increase sample sizes for rare disorders. Bayesian models can "learn" from external data on a per-feature basis, enabling incorporation of external knowledge such as *Dahlin's Bone Tumors*, a reference of anatomic and demographic statistics of more than 10,000 bone tumors based on the Mayo Clinic registry [24]. This is the reason we designed our NBM to scale—it can add more training samples and diagnostic classes dynamically in real time. We are currently improving the system to be able to "learn" from external knowledge (Dahlin, PubMed studies, etc.) so that each attribute can be more reflective of true disease prevalence, rather than reflective of the knowledge that is encoded in only our collection, potentially improving tumor prediction.

Compliance with Ethical Standards Institutional review board approval was obtained. The requirement for informed consent was waived as this was a retrospective review of de-identified radiologic images with only age, gender, and brief clinical notes/diagnosis.

Appendix

Table 3 Detailed list of 18 features and all possible values in Naïve Bayes model

Feature	Values
Age	Decade bins: 0–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, 90–100+
Gender	Male, female
Number of lesions	Solitary, multiple
Bone location	Carpals, clavicle, femur, fibula, foot, hand, humerus, iliac bone, ischium, mandible, patella, pubis, radius, rib, sacrum, scapula, skull, sternum, tarsals, tibia, ulna, vertebrae
Longitudinal location	Apophysis, diaphysis, epiphysis, metadiaphysis, metaphysis, n/a
Proximal vs. distal	Proximal, middle, distal, not applicable
Transverse location	Medullary cavity, endosteum, cortex, periosteum, sessile, pedunculated, juxtacortical, soft tissue
Distribution	Central, eccentric, n/a
Density	Normal, ground glass, lytic, sclerotic, mixed lytic, and sclerotic
Matrix/texture	Normal, bone forming or osteoid, chondroid, septated, coarse trabeculae, central calcification
Transition zone/border	Geographic 1A (narrow sclerotic), geographic 1B (narrow non-sclerotic), geographic 1C (wide non-sclerotic), permeative/destructive/punched out, unable to determine border or n/a (e.g., osteochondroma does not have a border)
Cortex	Endosteal scalloping grade: 0 = none, 1 = 0–25%, 2 = 25–50%, 3 = 50–75%, 4 = 75%+, where % = approximate depth of scalloping; cortical thickening, periosteal scalloping (any degree), n/a
Periosteum	No periosteal reaction, solid periosteal reaction, lamellated periosteal reaction, interrupted periosteal reaction, codman triangle, sunburst
Lesion to shaft ratio	0–25%, 25–50%, 50–75%, 75–100%, >100%, n/a
Physis	Closed, open
Expansion	Non-expansile, expansile
Soft tissue mass	Yes, no
Pathologic fracture	Yes, no

Table 4 Raw computations that are generated by the NBM in real time for any unknown query

Feature	Chondroblastoma	Giant cell tumor	Ganglion cyst
Age bin	0.28	0.36	0.22
Gender	0.61	0.51	0.44
Number of lesions	1.00	1.00	0.88
Bone location	0.14	0.34	0.11
Longitudinal location	0.38	0.04	0.22
Proximal vs. distal	0.14	0.58	0.44
Transverse location	0.95	0.97	0.77
Distribution	0.76	0.58	0.66
Density	0.57	0.87	0.88
Matrix/texture	0.47	0.68	0.77
Transition zone/border	0.57	0.75	0.11
Cortex	0.57	0.95	0.44
Periosteum	0.95	0.75	1.00
Lesion to shaft ratio	0.28	0.07	0.11
Physis	0.76	1.00	0.88
Expansion	1.00	0.48	0.88
Soft tissue mass	0.95	0.90	1.00
Pathologic fracture	0.90	0.70	1.00
Diagnosis probability	50%	48%	2%

When a query is submitted, the Naïve Bayes machine computes prior probabilities for all attributes given the query instances for each feature. The product of the prior probabilities is factored and then normalized to generate an overall class probability, or probability of diagnosis for *all* diagnoses. The diagnoses with conditional probabilities >0% are displayed in descending order as a "Ranked Differential Diagnosis." In this example, demographic and radiographic features were entered for an unknown lytic lesion in the distal diaphysis of the femur in a 25-year-old male patient. The prior probabilities for each of the 18 attributes and overall *normalized* diagnosis probability are calculated in real time. The top three "differential" are shown below with raw prior probabilities. This data is normally suppressed to the user

References

1. Costelloe CM, Madewell JE: Radiography in the initial diagnosis of primary bone tumors. AJR Am J Roentgenol 200(1):3–7, 2013
2. Rajiah P, Ilaslan H, Sundaram M: Imaging of primary malignant bone tumors (nonhematological). Radiol Clin N Am 49(6):1135–1161, 2011 **v**
3. Burnside ES: Bayesian networks: computer-assisted diagnosis support in radiology. Acad Radiol 12(4):422–430, 2005
4. Liu YI et al.: A bayesian network for differentiating benign from malignant thyroid nodules using sonographic and demographic features. AJR Am J Roentgenol 196(5):W598–W605, 2011
5. Lodwick GS et al.: Computer diagnosis of primary bone tumors: a preliminary report. Radiology 2(80):3, 1963
6. Lodwick GS: A probabilistic approach to the diagnosis of bone tumors. Radiol Clin N Am 3(3):487–497, 1965
7. Reinus WR et al.: Diagnosis of focal bone lesions using neural networks. Investig Radiol 29(6):606–611, 1994
8. Kahn, Jr CE, Laur JJ, Carrera GF: A Bayesian network for diagnosis of primary bone tumors. J Digit Imaging 14(2 Suppl 1):56–57, 2001
9. Richardson ML: Bayesian bone tumor diagnosis. 2005; Available from: http://uwmsk.org/bayes/bonetumor.html
10. Doyle LA: Sarcoma classification: an update based on the 2013 World Health Organization Classification of Tumors of Soft Tissue and Bone. Cancer 120(12):1763–1774, 2014
11. Rubin DL et al.: iPad: Semantic annotation and markup of radiological images. AMIA Annu Symp Proc:626–630, 2008
12. Murphey MD et al.: Enchondroma versus chondrosarcoma in the appendicular skeleton: differentiating features. Radiographics 18(5):1213–1237, 1998 quiz 1244-5
13. Oudenhoven LF et al.: Accuracy of radiography in grading and tissue-specific diagnosis—a study of 200 consecutive bone tumors of the hand. Skelet Radiol 35(2):78–87, 2006
14. Miller TT: Bone tumors and tumorlike conditions: analysis with conventional radiography. Radiology 246(3):662–674, 2008
15. Cerase A, Priolo F: Skeletal benign bone-forming lesions. Eur J Radiol 27(Suppl 1):S91–S97, 1998
16. Atesok KI et al.: Osteoid osteoma and osteoblastoma. J Am Acad Orthop Surg 19(11):678–689, 2011
17. Langlotz CP: RadLex: a new method for indexing online educational materials. Radiographics 26(6):1595–1597, 2006
18. Huynh BQ, Li H, Giger ML: Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imaging (Bellingham) 3(3):034501, 2016
19. Rajkomar A et al.: High-throughput classification of radiographs using deep convolutional neural networks. J Digit Imaging 30(1):95–101, 2017
20. Kooi T et al.: Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 35:303–312, 2017
21. Wang J et al.: Discrimination of breast cancer with microcalcifications on mammography by deep learning. Sci Rep 6:27327, 2016
22. Lejbkowicz I et al.: Bone browser a decision-aid for the radiological diagnosis of bone tumors. Comput Methods Prog Biomed 67(2):137–154, 2002
23. Rokach L, Maimon O: Data mining with decision trees: theory and applications. In: Series in Machine Perception and Artificial Intelligence, 2nd edition. Singapore: World Scientific Publishing Company, 2014, p. 328
24. Unni KK, Inwards CY: Dahlin's Bone Tumors: General Aspects and Data on 10,165 Cases. Philadelphia: LWW, 2010, p. 416